Analysis of chemotherapy screening assays using MCMC methods

Hollyer, Justine

Award date:
2000

Awarding institution:
University of Bath

Analysis of Chemotherapy Screening Assays using MCMC Methods

submitted by

Justine Hollyer

for the degree of Ph.D

of the

University of Bath

2000

ProQuest LLC
789 East Eisenhower Parkway
P.O. Box 1346
Ann Arbor, MI 48106-1346

Abstract

Aims

Chemotherapy screening assays are used to try and identify which drugs induce a response in patients *ex vivo* in the hope that it will correlate well with an *in vivo* response. This thesis is primarily concerned with identifying the dose at which 90% kill rate is achieved (the LC90) and comparing these results to get an indication of which patients are likely to respond to which drugs.

Data

The finalised data set consisted of over 40,000 assay results on 42 different drugs. Each patient had several assays performed on a number of these drugs. The number of drugs tested varied across patients depending on drug availability and previous treatment history.

Assay Methods

Slides are prepared with 5 "spots" of untreated cells. Control slides are left untreated to provide a baseline count whilst the other slides are treated with 5 decreasing dose levels for each drug tested. Cell survival is noted at each dose and recorded in categories which represent overlapping proportions of cell survival.

Statistical Analysis

We develop a model using all the prior information we have. The Metropolis-Hastings algorithm is used to construct a Markov Chain and an analysis run for each drug. From these analyses we note the posterior distributions for all variables and report estimates of the slope and location of the dose response curve as well as the LC90s.

We use the information gained about the posterior distributions of our variables in order to develop a further model to analyse extra cases without re-running the whole analysis.

Lastly we consider the problem of outliers and we look at methods to identify them and assess their influence.

Results

The resulting LC90s have great variation both between drugs and within drugs.

The accuracy of the estimate depends greatly on the quality of data. Small control values and missing data contribute to poor estimates.

An efficient procedure can be derived for looking at additional assays using the information gained from the initial main analyses.

The presence of outlying observations does not have a large effect on resulting estimates.

Convergence diagnostics often give conflicting results on burn-in, run length required and convergence depending on the technique employed.

Conclusions

We are confident that with sufficient data our estimates are reliable. In those cases where there is only a small number of cells to assess, these should be counted accurately rather than categorised into overlapping and often very wide bands. This will also reduce possible outliers.

We note that commonly used convergence diagnostics cannot wholly be relied upon. Work is currently active in this area.

The clinical implications from these results mean patients can have non invasive tests rapidly performed on a variety of drugs in order to obtain the best possible treatment. Drug induced resistance can be minimised by using these indicated drugs. Other benefits include cost and reduced side effects if more than one drug is shown to produce equal efficacy *ex vivo*.

Acknowledgements

My biggest thanks and a huge debt of gratitude go to my supervisor Professor Jennison whose encouragement, support and guidance never wavered.

I deeply indebted to my sisters who have provided both financial and emotional support as well as flowers for the many injuries I have sustained during my time at Bath!

Thanks are also due to the other members of the Statistics group at the University who have been willing to share their knowledge and ideas throughout both my MSc. and PhD. A special mention is also due to my various office mates for their occasionally stimulating conversations, but most often irrelevant and entertaining ones.

I am very grateful to the Royal Statistical Society for the generous use of their computing facilities during the time spent in London.

I acknowledge the EPSRC for their financial support.

Contents

List of Tables

List of Figures

Chapter 1

Introduction

The initial aim of this thesis is to develop a procedure which will utilise a data set of assay results obtained from patients with chronic lymphocytic leukaemia and hence produce accurate estimates of the effective dose required for treatment. The assays used are concerned with ascertaining the proportions of cancerous cells killed by adding different drugs to tumour samples at various concentrations.

1.1 Background

Information has been collected over a number of years by the Bath Cancer Research Unit to test the efficacy of a wide variety of drugs on tumour and blood samples received for analysis. Even though suffering from seemingly identical diseases, individuals respond very differently to different drugs making this a particularly difficult illness to treat. One consequence of failing to identify an effective treatment initially is a drug induced resistance which decreases the therapeutic effect of further drug regimes making subsequent treatment increasingly more difficult. It is known that patients who demonstrate sensitivity to particular drugs *ex vivo* will also often respond well *in vivo* and hence any potentially effective drugs indicated can be used in treatment. *Ex vivo* is used to denote work with fresh cells as opposed to cell lines *in vitro*. The objective of the assay therefore is to determine how likely it is that a patient will respond

to a particular drug by noting that the results obtained *ex vivo* correlate well with actual responses (Bosanquet, 1991 and 1995). Also, if previous standard therapies have already been shown to have failed, the assays can often identify drugs which may not have been previously considered and these will include "unusual" non-standard treatments.

The assay in use has been refined so it now has the ability to test a wide variety of drugs quickly and easily so results can be obtained with minimum delay. A patient is said to be sensitive to a drug if the dose at which at least 90% of cancerous cells are killed in the assay is low when compared to the results from other patients with the same disease. This does not necessarily imply an effective treatment however, merely that a patient has shown a better response when compared to results from other patients. Since many of these drugs are highly toxic with undesirable side effects, even seemingly low doses may not be tolerated.

The data set under investigation is large with over 40,000 records and is structurally complicated with many missing observations. Each record varies both in the amount of information held and its quality.

Every patient has had several assays performed from a selection of 40 drugs. Assays are duplicated to test for repeatability and also to provide a back-up measure in case of assay failure. The results may be recorded twice by different scientists to ensure agreement.

Owing to the structure of this data set and the need to estimate unobserved variables, Markov Chain Monte Carlo (MCMC) methods have been used since they offer a framework in which to simulate the actual processes involved in obtaining the data and hence produce estimates of unobserved variables where the main interest lies.

MCMC methods have become an increasingly common way in which to analyse data where more traditional methods are unable to cope with model complexity and intricate data structures. With ever developing computing power, such computational methods have become increasingly accessible on even small platforms. Whilst existing packages have been written utilising MCMC methods, they are limited in their use owing to their inability to analyse more complicated

models satisfactorily. Hence a program has been specially written as part of this project to cope with the data structures arising and provide estimates of the variables of interest.

The first part of this thesis is concerned with identifying a model which mimics the data collection procedure. This initially entailed developing the model using simulated data which had simpler structure than the final data set in use. The final data set proved to be far more complex than originally envisaged and meant extending the original model. Even using the simplified data set, its structure meant it was impossible to use standard analyses and packages satisfactorily. Since the results were recorded with varying amounts of accuracy, this also has to be accounted for. The missing data variables were simulated by running a Markov chain. Since non-standard distributions had to be used, a Metropolis-Hastings Algorithm was the basis for performing updates on the simulations.

Further aims of this project include the assessment of outliers in the data and their effect in the analyses. Also of importance is the ability to analyse new cases separately, without the need to run a full simulation each time. This is done by taking information from previous runs about the posterior distributions of the salient variables in order to provide information about each particular drugs' behaviour.

1.2 Structure of the thesis

Chapter 2: Medical Background

The next chapter is concerned with explaining the medical background and techniques employed to perform the assay and ensure results are as accurate as possible.

This chapter also demonstrates the usefulness in utilising these assay results. Those patients whose treatment is based on those drugs shown to be effective *ex vivo* have a considerably better prognosis than those who show no sensitivity to any of the drugs. It is hoped that by using the indicated drugs, survival times

of patients showing some sensitivity will improve even if treated with different drugs than those shown to have the best *ex vivo* results.

Chapter 3: Bayesian Analysis using MCMC

This chapter gives a brief introduction to Bayesian methods and the underlying theory. It deals with the two most popular methods of implementation of Markov Chain Monte Carlo and surrounding issues.

Chapter 4: Model Specification

We describe the model in use and the stages in its development. The initial data set used to test the initial model was simulated to be a smaller and simplified version of the one in final use. As the structural complexities of the final data set were introduced, the model evolved to cope with these features. The choice of prior distributions for the variables specified in the model was based on information gathered from the scientists. These were chosen to include all possible values which may arise so whilst they were not non-informative, they were sufficiently diffuse to cope with all potential values which were thought to be possible within the data.

Chapter 5: Model Implementation

Here we discuss the choice of algorithms used to implement the model described in Chapter 4. This breaks down into four major procedures: reading in of data into suitable data structures, initialisation of all the variables to be updated, the updating of these values whilst running the Markov chain and finally the output of the results. We take particular care with specifying the acceptance probability calculations since they are vital in obtaining suitable posterior distributions.

Chapter 6: Model Validation

In order to satisfy ourselves that the results from the model are reliable we perform validation checks on the output. These consist mainly of convergence diagnostics which have been implemented in a public package called CODA. This chapter looks at the results of various techniques and compares the diagnostics of the different methods.

Chapter 7: Results

Although we are primarily interested in the estimates of the LC90s and how reliable they are, we are also interested in the resulting cell survival curves and the associated parameters. We look at the variation between drugs, between patients within a particular drug and also how the different assays compare for the same patient. We also consider the problem of examining an additional patient after the main analysis has been carried out.

Chapter 8: Outliers

This chapter is concerned with the effect of potential outliers. It looks at ways of identifying them and various theories about their treatment. We deal with them by omitting the suspect values and treating them as though they were missing. The resulting posterior distributions for the observations in question are compared with the omitted values in order to compare their discordancy and also to assess the possible influence on the estimated LC90 and cell survival curves.

Chapter 9: Conclusions

Finally we present a review of the results obtained and the methods used in their analysis. We look at possible recommendations for implementing new data collection procedures and further work which may be carried out in the future.

Chapter 2

Medical Background

The aim of cancer therapy is to kill tumour cells and/or prevent division of cancer cells and hence arrest the disease. There is good evidence to indicate that assays performed on patient cell samples which measure either *in vitro* or *ex vivo* the kill rate of cells or the inhibition of cell proliferation may be indicative of clinical response. (*In vitro* refers to cell lines and *ex vivo* denotes work with fresh cells whereas the term *in vivo* refers directly to the patient). *In vitro* drug sensitivity assays have been used for over 40 years as a way to identify potential *in vivo* sensitivity of individual patients to specific anti-cancer chemotherapy (Bosanquet 1994). *In vivo* denotes the actual patient response. Even patients with seemingly identical diseases can respond very differently to the same drug, which makes it very hard to prescribe effective treatment prospectively. Hence, many *in vitro* drug sensitivity tests have been developed with much work being concentrated on short-term assays and their predictive accuracy, i.e. the agreement between test result and actual patient response. Bosanquet (1991) demonstrated that patients with a low tumour cell survival *in vitro* (i.e.,less than the median) have an improved response rate *in vivo* when compared to the overall response. This improvement is almost double that of the overall response rate for the particular drug and disease.

2.1 Assays

Older assays used in the 1970s took 2 to 3 weeks to perform and were labour intensive, consequently their use was limited. Since then, many shorter term tests taking less than a week have been developed, making them far more convenient to use. One such assay is the Differential Staining Cytoxicity assay (DiSC) originally devised by Weisenthal (Weisenthal, Marsden, *et al.* (1983), Weisenthal, Dill, *et al.* (1983)) taking around 4 days. This DiSC assay has been further developed to make it particularly well suited to the testing of large numbers of drugs using relatively small tumour samples (Bird *et al.* 1988). It has therefore been the method used to test all the samples examined in this study.

At the Bath Cancer Research Unit in the Royal United Hospital, Bath, samples of both solid tumours and leukaemias are analysed in an effort to identify the drugs which are most likely to produce a response in the patient. This information can then be referred back to the physician who is thus empowered to make an informed choice about the treatment of a patient by selecting a regime from those drugs shown to be effective.

2.1.1 Benefits

Apart from the obvious need to find a cure or arrest the spread of disease, identifying the best potential treatment regime indicated from these assays has further implications. There are many drugs available and patient response differs widely to each treatment. Such testing enables an informed choice of chemotherapy to be made on the basis of a response *ex vivo*. Patients experience drug induced resistance: the more courses of chemotherapy received, the less likely the patient is to respond to further courses. If an ineffective treatment is given initially, it will reduce the beneficial effect of a more efficacious treatment given later on. Hence the therapeutic benefit of a potentially effective treatment is reduced if previous courses have already been administered. Combinations of some drugs, especially those which work similarly, have high drug cross-resistance patterns to a response *ex vivo* (Bosanquet 1991).

Also, many of these drug have high toxicities with unpleasant side-effects. So, if two drugs give similar results *ex vivo* but one is known to have less severe side effects this could be given in preference. Similarly with cost, some treatments, especially those which are platinum based, can be extremely expensive so patients could be given a cheaper drug without detriment.

The DiSC assay thus enables many drugs to be tested simultaneously on tumour samples, enabling the physician to identify potentially effective treatments which may not have been otherwise considered. This is especially helpful if previous courses of standard treatment have already failed. If used beforehand the assays may in fact have shown that these standard drugs are likely to have little beneficial effect so they can be avoided altogether.

2.1.2 Clinical correlations

In order for the assay results to be reliable guides for treatment choice, *ex vivo* results should agree well with *in vivo* results. A clinical correlation is obtained by categorising both the *ex vivo* and *in vivo* responses into either sensitive or resistant cases and then comparing the agreement of these variables. A patient response is defined as either a complete or partial response. However, an assay response is not so easily dichotomised since results cover a broad range from very sensitive to very resistant. Picking a cut-off point can only be done satisfactorily with reference to the clinical data.

2.2 Assay methods

For each patient, a number of slides each with five "spots" of the sample tumour are prepared. (Figure 2-1).

Two control slides are left untreated in order to obtain a "baseline" count for the number of cells examined. Up to 40 further slides are treated with different drugs. Each slide is treated with a single drug. The drugs themselves are each diluted four times by a constant factor to give five different concentrations. The

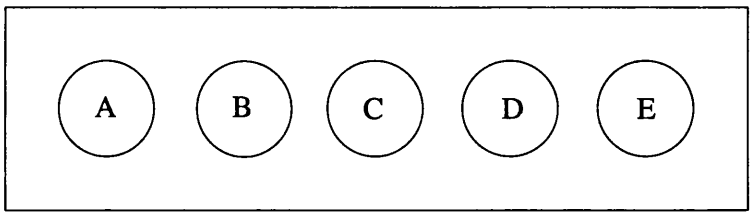

Figure 2-1: *Diagrammatic representation of treated slide. Spot "A" is treated with the maximum concentration (MC), Spot "B" is treated with MC/DF, where DF= the dilution factor, "C", "D" and "E" are treated with* MC/DF^2, MC/DF^3 *and* MC/DF^4 *respectively.*

dilution factor used depends on prior knowledge of the *ex vivo* performance of that particular drug and will hopefully span the cell survival rate of interest. Even for the same drug, different dilution factors may be used as knowledge of a particular drug's behaviour is gained. The drugs are incubated with cells for 4 days in a controlled environment. A known number of duck blood cells are added as a control measure: the number of tumour cells is not known in advance. The cells are then stained and then centrifuged onto collagen coated slides. The slides are fixed and counterstained before examination. Duplicate back up slides are also made as a precaution against failure or contamination. These duplicates are sometimes used as an additional measure to ensure consistency.

Up to two scientists are employed in counting the baseline number of cells and hence ascertaining the proportion of cells killed. So for each patient there is a maximum of four sets of results for each drug tested. The control and treated slides are stored in the same box and analysed together.

2.2.1 Control measures

One of the problems during preparation of these slides comes when they are centrifuged: it is not unknown for cells to "disappear". In order to assess any cell loss, a known quantity of duck cells are added. If all the duck cells are still present after the centrifusion then it can be safely assumed that none of the patients cells have been lost either and the slide is reliable. The different types of cells are easily distinguishable when examined under a microscope: after staining, the duck cells appear as large blue "blobs" whereas the patient cells are far smaller and range

in colour from bright magenta for live cells to nearly black for dead and dying ones. Thus it is relatively straightforward to differentiate the live and dead cells from the duck cells and ascertain the reliability of the slide. If a slide is seen to have failed, the backup slides are used instead.

Cross contamination can also occur with some drugs during incubation. Certain drugs not only react with the the cells to which they have been added, but also to others in the same box leading to higher kill rates than would occur normally. Care is taken to ensure these drugs are stored and incubated separately.

2.2.2 Counting methods

In order to get a control value, i.e., the expected number of live untreated cells for each patient per square, the cells are counted using a microscope with a 10×10 grid of squares in the eyepiece. The lens is positioned over a "representative" area of each spot in turn on the first control slide. The number of cells in one grid square are counted manually and the value noted. Owing to the irregular distribution of the cells, 10 such squares are chosen to be counted and then averaged to give a baseline count.

The second control slide is counted in the same manner but unless the resultant baseline value differs from the first (by more than 30%), only the first value is recorded.

The number of treated cells is ascertained in a similar manner but the absolute value of remaining cells is not noted, instead a number of overlapping categories is used to record this information. These categories are very narrow where there is a low survival rate, since this is of most interest, and conversely, where the survival rate is high, the categories have considerable overlap and are relatively wide. The reason for recording data in this manner is the considerable reduction in time spent counting individual cells where accuracy is not imperative. Extra care is taken where survival rates are low and actual counts are made, otherwise the proportion surviving relative to the control value is estimated. Owing to the random distribution of cells it is possible to obtain estimated kill rates which are greater than 100% if the number of dead cells on the treated slides exceeds that

	Category name	Proportion of cells surviving
1)	0	0 - 2%
2)	5	0.1 - 8%
3)	10	2 - 12%
4)	15	10 - 25%
5)	30	15 - 45%
6)	45	30 - 60%
7)	60	45 - 100%
8)	100	> 60%

Table 2.1: *Initial categories used to denote proportion of surviving cells*

of the control slide. This is more likely when there is only a small quantity of control cells.

Since the first recording of results, more categories have been introduced to make the recording of data more accurate and the ones in use at present are shown in Table 2.1

2.2.3 LC90

The quantity of interest is the *Lethal Count* 90 (LC90): that is the drug concentration at which 90% of cells are killed. Since only 5 drug concentrations are used this was initially estimated as the weakest concentration at which less than 10% cells survive. If the sample was very resistant, the LC90 would often not be achieved at the concentrations tested. The converse problem could also happen when the sample is very sensitive and even the lowest dose would kill all cells.

The maximum concentration and dilution factors were chosen to try and span the LC90 but, given the vast range of responses, this was not always possible. As further information was gained about particular drug behaviour, the dose levels were changed to reflect this additional knowledge.

Given the nature of the data, the estimate of the LC90 is subject to considerable

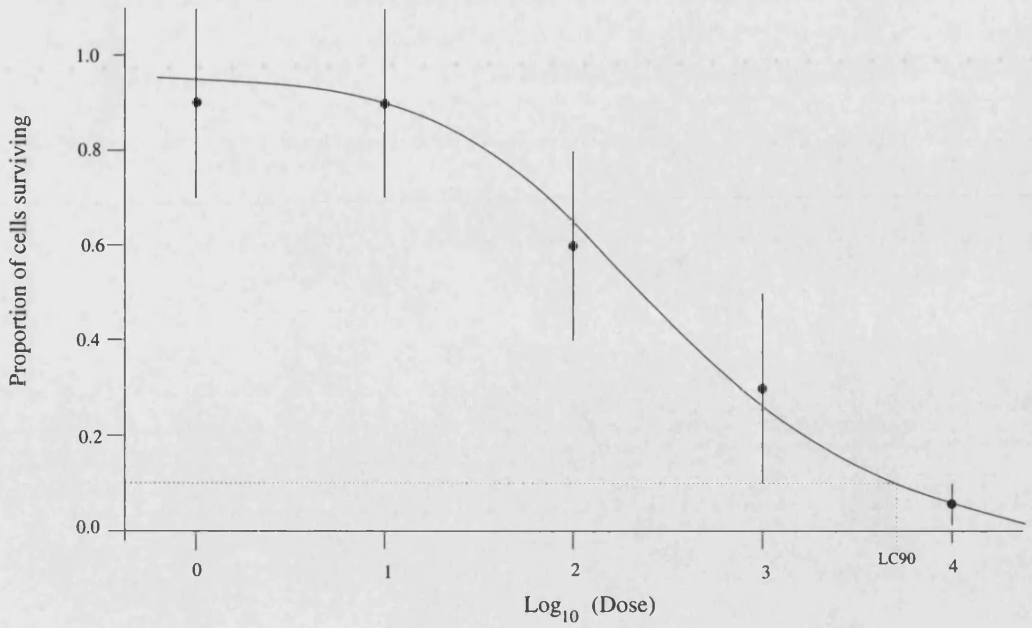

Figure 2-2: *Estimation of the* LC90. *The vertical bars represent the estimated range of surviving cells. The solid line denotes the curve fitted to the mid-point of these ranges. The dotted line touches the curve at the point where 90% of cells are estimated to be dead and the relevant concentration (LC90) is read from the x-axis.*

error and part of the aim of this project is to give more accurate estimates. The LC90 has previously been ascertained by graphing the midpoint of the interval of the cells surviving, fitting a curve to these 5 sets of points and then reading off from the axis the drug concentration corresponding to 10% survival (Figure 2-2).

2.2.4 Sensitivity

A patient is said to be sensitive to the drug *ex vivo* if the LC90 is achieved at a low dose when compared to other patients with the same disease. However this is subject to misinterpretation. A patient may have a good result when compared to other patients but the drug may, in general, be very poor at killing the cancerous cells. So a "good" assay result does not necessarily mean the patient will respond well *in vivo* and this has to be taken into consideration when reporting the results back to the doctors.

2.3 Survival analyses

It has been shown that response rates and length of survival can be considerably enhanced utilising the results from these assays. Even allowing for known effects such as stage of disease, age and number of previous courses, the patient sensitivity to drugs tested *ex vivo*, whether exploited or not, is also an important independent factor in the patient's response and overall survival. A subset of data with further information regarding survival times and extra prognostic factors was analysed. The resulting Kaplan-Meier curves can be seen in Figure 2-3 and they do confirm these assertions.

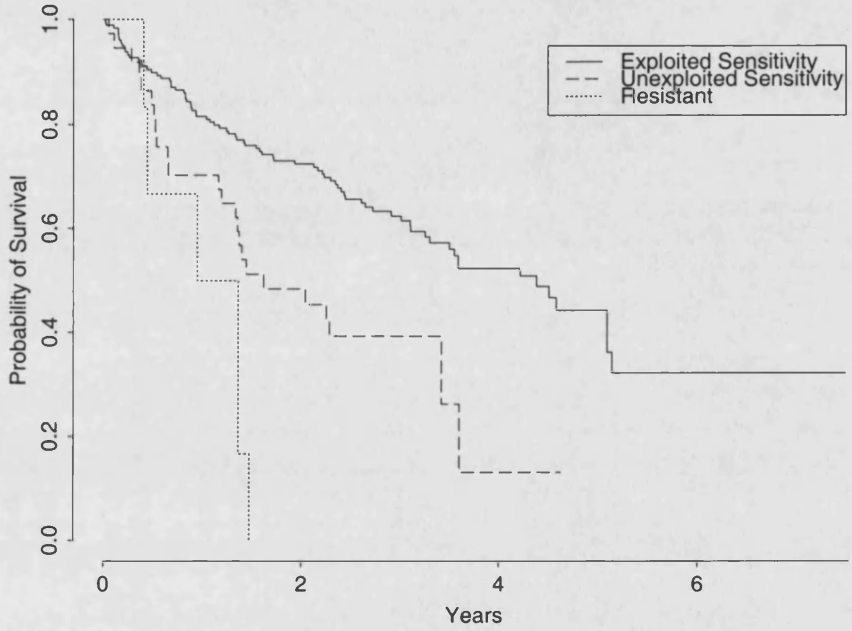

Figure 2-3: *Kaplan-Meier Curves stratified by* ex vivo *sensitivity. The resistant group (bottom curve) have the worst estimated median survival at 1.0 years, the unexploited and exploited groups have median survival of 1.5 and 4.2 years respectively.*

Those patients whose treatment includes any of the drugs shown to be effective *ex vivo* from the assays (the *exploited sensitivity* group) have the best survival rates whereas those shown not to be sensitive have, predictably, the worst. The intermediate group are those patients who have been shown to have a response

ex vivo to at least one of the drugs tested but for various reasons none of the indicated drugs have been used in the treatment prescribed. This *unexploited sensitivity* group, if treated using those drugs to which the patient is likely to be sensitive, are likely to have a survival curve approaching that of the *exploited sensitivity* group.

It should be noted that the data on which this particular analysis is made are observational rather than from a randomised clinical trial. However, the differences in survival between these groups are very marked and extremely unlikely to be due to chance or any possible bias.

Thus there is compelling evidence that assay-based treatment does lead to better survival and remission rates than those observed for patients with unexploited sensitivity. Both these groups out perform the assay-resistant patients for whom the only realistic treatment is palliative or investigational.

Chapter 3

Bayesian Analysis using MCMC

MCMC methods provide a framework which enables complex statistical modelling to be carried out via simulations. Traditional methods could require either a simplification of the problem in order to utilise existing methods or new methodologies which may have to be developed.

3.1 An introduction to Bayesian analysis

We shall use the principles of Bayesian data analysis to build models and summarise unobserved quantities of interest. In order to make inferences on variables we cannot examine we use information from data which we can observe. This is done by obtaining a posterior distribution for the variable of interest. A full probability model is specified in which all observable and unobservable quantities are treated as random variables, each with associated prior distributions. Bayesian modelling does not distinguish between observables and parameters of a statistical model: they are all considered to be random quantities with individual distributions. The prior probability distribution must be consistent with knowledge about the underlying problem and the likelihood with data collection processes, for example whether the data are discrete, or subject to a restricted range.

The joint distribution can be specified over all random quantities, which comprise

two parts: a *prior* distribution $P(\theta)$ and a *likelihood* $P(D|\theta)$ where θ denotes the model parameters and D the data observed.

The joint probability model can then be specified in terms of the prior distribution and the likelihood,

$$P(D, \theta) = P(D|\theta)P(\theta). \tag{3.1}$$

Bayes theorem is used to obtain the *posterior* distribution of θ conditional on the data D,

$$P(\theta|D) = \frac{P(\theta)P(D|\theta)}{P(D)}, \tag{3.2}$$

where $P(D) = \int P(\theta)P(D|\theta)d\theta$.

Since $P(D)$ does not depend upon θ, it may be omitted yielding the *unnormalised posterior density*:

$$P(\theta|D) \propto P(\theta)P(D|\theta). \tag{3.3}$$

Bayes Rule means that the data D affect the posterior inference only through $P(D|\theta)$ - the likelihood function. In this way, Bayesian inference obeys the *likelihood principle*, which states that for a given sample of data, any two probability models $P(D|\theta)$ that have the same likelihood function yield the same inference for θ (Gelman *et al.* 1995). Bayesian methods enable statements to be made about the partial knowledge available, based on data, concerning some situation (unobservable or as yet unobserved) in a systematic way in agreement with probability theory. The guiding principle here is that the state of knowledge about anything unknown is described by a probability distribution.

Thus the task is to specify the full model $P(\theta, D)$ and compute summaries for $P(\theta|D)$. With MCMC methods this is achieved by sampling from a Markov chain which has been constructed to have the correct limiting distribution. Inferences can then be made from summaries which can be obtained from these samples.

3.2 Markov Chain Monte Carlo sampling

Suppose we wish to calculate $E[f(X)]$ for a specified function f when X follows the distribution π. If it were possible to draw independent samples $\{X_i; i = 1, ..., n\}$ from the desired distribution $\pi(x)$ we could approximate $E[f(X)]$ by the sample mean

$$\frac{1}{n}\sum_{i=1}^{n} f(X_i). \tag{3.4}$$

However, if π is a complex distribution, it may be infeasible to obtain such samples directly.

The idea in MCMC sampling is to create a Markov Chain with *limiting* distribution equal to π, the distribution of interest, and treat samples obtained after the Markov chain has run for a long time as realisations from this distribution. For an aperiodic Markov chain, the stationary distribution is also the limiting distribution. If $\{X_i : i = 1, 2, ...\}$ form a Markov chain, X_{i+1} depends on $X_0, ..., X_i$ only through X_i and under certain conditions the chain gradually "forgets" its initial state and the distribution of X_n converges to the limiting distribution as $n \to \infty$.

In order to construct such a chain we need to carry out the following procedures

- Set up a chain with a limiting distribution of π.
- Discard the burn-in, the initial m samples.
- $E[f(X)]$ can then be estimated by

$$\frac{1}{n-m}\sum_{i=m+1}^{n} f(X_i).$$

After a sufficiently long initial burn-in period of m iterations, $X_{m+1}, ..., X_n$ will be dependent samples from the stationary distribution $\pi(.)$. Realisations from such a chain are necessarily correlated but the $\{X_i\}$ do not necessarily need to be independent in order to be useful.

There are, however, certain conditions that the Markov chain must satisfy in order for it to produce reliable samples (Tierney, 1995).

- Irreducibility: The chain must be able to reach all parts of the state space from any other part of the state space.
- Recurrence: This is the property that all parts of the the chain will be reached infinitely often, at least from almost all starting points.
- Convergence: The chain must first achieve convergence to a stationary distribution in order for samples to have the required posterior distributions. There are numerous methods by which to assess convergence which are discussed in further detail later.

The X_i are generated such that X_{i+1} depends only on X_i. Hence it depends only on X_0 through the intervening variables $(X_1, X_2, ..., X_{i-1})$. Thus the chain gradually forgets its initial state and will eventually converge to a unique stationary distribution.

Convergence to the required distribution is ensured by the *ergodic theorem* which states

$$P(X_{k+N} = j | X_k = i) \to \pi_j \quad \text{as } N \to \infty \tag{3.5}$$

where $\pi = (\pi_1, ..., \pi_n)$ is the unique solution to $\pi P = \pi$, (if π is a discrete distribution).

Burn in samples are discarded giving the parameter estimator as the *ergodic average*:

$$\overline{f} = \frac{1}{n-m} \sum_{i=m+1}^{n} f(X_i). \tag{3.6}$$

3.3 Two implementations of MCMC

3.3.1 The Metropolis-Hastings Algorithm

Hastings (1970) further developed the sampling method introduced by Metropolis *et al.*(1953) into a more generalised form. In order to obtain these samples, the Metropolis-Hastings Algorithm is thus used to construct a Markov Chain with its stationary distribution equivalent to the one of interest.

To construct such a chain the algorithm consists of the following steps:

1. Initialize $X_0 = x_0$
2. At time i, when $X_{i-1} = x_{i-1}$, for x_i, generate proposal value x' for X_i.
3. Let $X_i = x'$ with probability

$$\alpha(x_{i-1}, x') = \min\left(1, \frac{\pi(x')q(x', x_{i-1})}{\pi(x_{i-1})q(x_{i-1}, x')}\right) \tag{3.7}$$

   otherwise $X_i = x_{i-1}$
4. Repeat from step 2 until desired length of chain achieved after burn-in.

From this it can be see that $\pi(x)$ does not need to be fully calculated since as ratios are used any normalising constant cancels, making the calculations considerably less complex.

For each state of X_i, the transition probabilities q_{ij} need to be defined such that the transition matrix Q is irreducible: every state for X_i can be reached from every other state X_j via intermediate states.

If this transition matrix is symmetric then the method simplifies to the Metropolis Algorithm.

The proposal distribution $q(.|.)$ can have any form and the stationary distribution will be $\pi(.)$. This can be seen by the following argument:

$$P(x_{i+1}|x_i) = q(x_{i+1}|x_i)\alpha(x_i, x_{i+1})$$

$$+I(x_{i+1} = x_i)\left[1 - \int q(x'|x_i)\alpha(x_i, x_i')dx'\right], \tag{3.8}$$

where $I(.)$ denotes the indicator function. The first term arises from the acceptance of proposal value $x_{i+1} = x'$ and the second from rejection for all possible values of x_{i+1}. From 3.7

$$\pi(x_i)q(x_{i+1}|x_i)\alpha(x_i, x_{i+1}) = \pi(x_{i+1})q(x_i|x_{i+1})\alpha(x_{i+1}, x_i). \tag{3.9}$$

Using Equations 3.8 and 3.9 for $X_{i+1} \neq X_i$ we then obtain the *detailed balance* equation:

$$\pi(X_i)P(X_{i+1}|X_i) = \pi(X_{i+1})P(X_i|X_{i+1}) \tag{3.10}$$

Integrating both sides with respect to X_i gives:

$$\int \pi(X_i)P(X_{i+1}|X_i)dX_i = \pi(X_{i+1}). \tag{3.11}$$

The left hand side gives the marginal distribution of X_{i+1} under the assumption that X_i is from $\pi(.)$. Therefore X_{i+1} is also from $\pi(.)$.

It can be seen that a starting value is required for X. This choice is not crucial since owing to the Markov property the chain eventually "forgets" where it started and as long as the chain is irreducible, the stationary distribution is not affected. However, very extreme values will adversely affect the length of the burn-in period and hence the time taken for the chain to converge.

Acceptance rates

Proposal distributions must be chosen carefully since small steps $(X_{i+1} - X_i)$ will have a high acceptance rate but mix slowly. Conversely, large steps will often propose moves to areas of low probability and these will be rarely accepted.

3.3.2 The Gibbs Sampler

Gibbs sampling (Geman & Geman 1984) is a special case of the Metropolis-Hastings algorithm. It is one of the most common ways in which to construct a Markov chain since it requires little more than sampling from the full conditional distributions (Gilks *et al.* 1996).

Conditional distributions are derived from the joint distribution of variables π. However, unless the prior distributions are conjugate to the likelihood, it may be infeasible to calculate the full conditional distributions.

Given a vector of unknowns, $u = (u_1, u_2, ..., u_K)$, let $u_{(i)}$ denote u with component i omitted, i.e. $u_{(i)} = (U1, ..., u_{i-1}, u_{i+1}, ..., u_K)$. Then with observed data D, the full conditional distribution for component u_i is:

$$(u_i | u_{(i)}, D).$$

One implementation of Gibbs Sampling is in the form of the package BUGS (Spiegelhalter *et al* 1994) which constructs a chain using prior distributions and conditional relationships as defined by the user. This will be looked at in fuller detail in a later chapter.

3.4 Burn in and convergence

It has been mentioned that, when running a Markov chain Monte Carlo, a burn in period is required in order for the chain to reach its stationary distribution on which our subsequent inferences may be based.

It is impossible to tell in advance when convergence will been reached and how long the burn-in period should be. However, although it is relatively simple to see whether a chain *hasn't* converged, the converse is unfortunately not true.

Many convergence diagnostics have been developed, some of which have been implemented in the package CODA (Best *et al*, 1995) which will be examined in

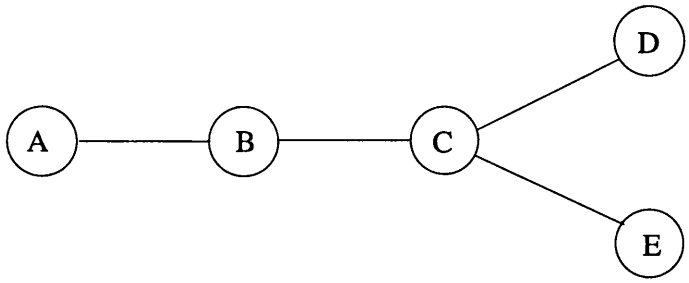

Figure 3-1: *Example of simple graph.*

more detail in a further chapter.

A simple and informative diagnostic is a plot of the chain. This will show whether the chain is mixing well as well as whether convergence may have been reached. This is not foolproof however, for example the posterior distribution may be multimodal and the chain may be sampling from just one mode with low probability of reaching another mode. In this case the chain will need to be run for longer in order for it to reach all areas of probability, but there may be no initial indication of this.

3.5 Graphical Models

If we wish to specify a model for random vector X with a complex multivariate distribution, it is useful to be able to express any association between the elements graphically. A convenient way of expressing such information is to construct a graph.

3.5.1 Graphs

A graph consists of a set of N nodes and E edges connecting the nodes. Each element of E is a pair of nodes (A, B) which denotes an edge between A and B. For example, the simple graph in Figure 3-1 show 5 nodes $\{A, B, C, D, E\}$ and 4 edges $\{(A, B), (B, C), (C, D), (C, E)\}$.

3.5.2 Conditional independence

A link between a pair of nodes indicates a relationship between variables and confers the status of *neighbours* between them. The absence of such a link between nodes denotes conditional independence given any intermediate nodes. The graph in Figure 3-1 shows the relationship between 5 variables. A is independent of all other nodes given B, written as: $A \perp\!\!\!\perp C, D, E|B$, where $\perp\!\!\!\perp$ denotes the independence operator. In the same way B is independent of D and E given C, denoted by: $B \perp\!\!\!\perp D, E|C$. In general, a node is independent of all other nodes given its immediate neighbours.

3.5.3 Directed graphs

In a *directed* graph, each edge $\{A, B\}$ is an ordered pair with a connection from A to B. We refer to the antecedents of a node as *offspring.* Each offspring will have a set of *parents* unless they are *founder nodes* with no parents.

Introducing directed links introduces the concept of ordering of the nodes. For a set of nodes $K = \{1, 2, ...k\}$ such that for all i and j in this set, either $i \prec j$ or $j \prec i$. We can then write $1 \prec 2 \prec .... \prec k$. In a directed graph, such ordering means that any edge can have only one possible direction.

3.5.4 Directed acyclic graphs

Extending graphs to include directed edges yields the problem of cycles. Initially this may at first seem desirable in that it would appear to have possibilities of modelling "feedback". Unfortunately there is no suitable joint probability to model such a situation, hence graphs used to model conditional dependence are not permitted to contain directed cycles.

It turns out that restricting the directed graph to exclude directed cycles, yielding a directed acyclic graph (DAG), is equivalent to supposing the nodes are completely ordered (Whittaker 1989).

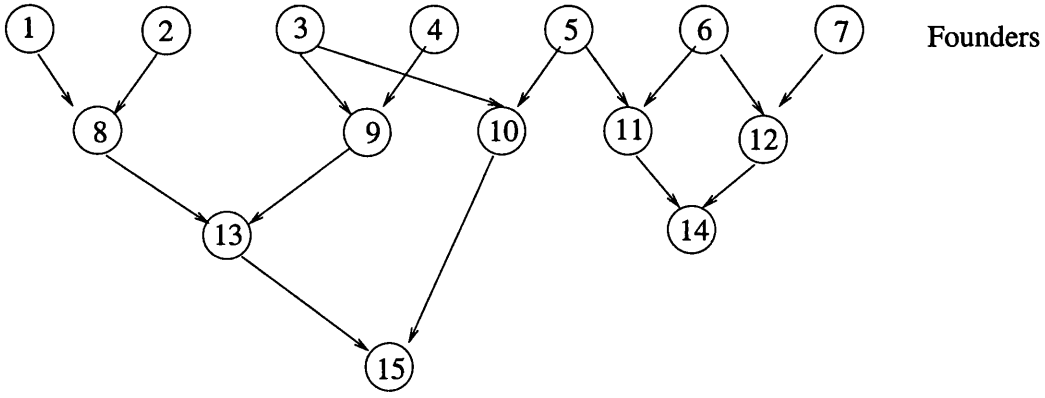

Figure 3-2: *Example of a directed acyclic graph (DAG) showing a Pedigree.*

Figure 3-2 shows a naturally occurring example of such a graph stemming from genetics and depicts a pedigree in which there are founders, parents and offspring. Nodes {1,2,..., 7} are the founders, nodes {8,..., 12} are the offspring of the founders etc. Whilst the nodes in this graph are necessarily compelled to have two parents owing to its data structure, this is not a general constraint and nodes may have any number of parents.

3.5.5 Graphical models

A graphical model is a probability model for multivariate random observations whose independence structure is characterised by a graph. It consists of a family of probability density functions that incorporates a specific set of conditional independence constraints listed in a DAG. As such it is a useful tool to represent the relationships between variables in our model and offers a convenient way to encode this information.

We use DAGs to graphically represent the model we wish to implement. Here the nodes represent the variables and the relationships/dependencies between them are denoted by the links.

The conditional independence assumption means that the full joint distribution of all quantities has a simple factorisation in terms of the conditional distribution of each node given its parents: this is easily determined from the DAG. Hence only the parent-child distributions are required in order to fully specify the model.

With Bayesian models, the joint distribution of the data and parameters is a product of many terms, each involving only a subset of the parameters. Thus a DAG is a convenient and simple way of expressing the complex conditional inter-dependencies which may exist between variables.

The joint distribution can be specified as a product of all terms. For any node k, denoting the remaining nodes by $K\backslash\{k\}$, the full distribution is thus:

$$f_{1,2,\ldots k} = f_{k|K\backslash\{k\}} f_{(k-1)|K\backslash\{k,(k-1)\}} f_{(k-2)|K\backslash\{k,(k-1),(k-2)\}} \cdots f_{2|1} f_1. \tag{3.12}$$

In order to obtain the full conditional distribution for node k, we have

$$\begin{aligned} P(k|K\backslash k) &\propto p(k, K\backslash k) \\ &\propto \text{terms in } p(K) \text{ containing } k \\ &= p(k|\text{ parents}[k]) \prod_{k \in \text{ parents}[l]} p(l|\text{ parents}[l]). \end{aligned} \tag{3.13}$$

Hence the conditional distributions for each k depends on the values of its parents, offspring and the offsprings' other co-parents, greatly reducing the number of terms involved.

We therefore use DAGs as a way of expressing the graphical model for our problem. We use the conditional independence assumptions to reduce the complexity of the model since many components have a simple factorisation given its parents. Each node represents a variable in out problem. This may be directly observed e.g. the control value or a stochastic node, i.e. one we wish to simulate e.g. the number of surviving cells.

Chapter 4

Model Specification

4.1 Description of data

In order to develop an adequate model for the problem, it is essential to describe the mechanism by which the data arise. This involves modelling the process by which the data are collected which includes both recorded data, as well as unobserved or unrecorded "hidden" variables.

The initial data set was a smaller simplified version of the one in final use. This data set did not allow for more than one control value or more than one replicate of the assay data categories. It was also assumed that maximum concentrations and dilution factors were kept constant for each drug (as the recording of the assays became more sophisticated and more knowledge was gained, these were often changed). It was considered sufficient to include just five categories in which to record the number of surviving cells.

We discussed the interpretation of these categories with Dr Bosanquet and concluded that they corresponded to the ranges of cell survival shown in Table 4.1. Here, proportions are relative to the number of cells recorded for the control slide.

The information recorded for patient i is as follows:

1. $N0_i$ The number of cells observed per unit area on the control slide.

	Category name	Proportion of cells surviving
1)	0	0 - 5.5%
2)	10	4.5 - 12%
3)	30	10 - 50%
4)	60	40 - 85%
5)	100	> 70%

Table 4.1: *Final categories used to denote proportion of surviving cells.*

2. x_j where $j = 1, ..., 5$. The dose levels: this is calculated from the maximum concentration used, (c), and the dilution factor, (d).

3. Y_{ij} where $j = 1, ..., 5$. The category indicating the proportion of surviving cells treated at dose level j. In order to obtain the value of Y_{ij}, we require the number of surviving cells to be estimated: this number is not recorded but is important when developing the model.

4.2 Initial model

We need to describe the relationship formally between the various data values in order to construct a suitable model.

In addition to the data actually recorded, there are several unobserved variables we also need to include in order to be able to make any useful inferences about them.

We shall treat the number of control cells, $N0_i$, per area on a slide as a Poisson count. The category of surviving cells, Y_{ij}, is derived from the proportion of surviving cells which depends on both the number of cells per area on a control slide and the number of surviving cells N_{ij} per area for each dose level. The number of surviving cells depends on the initial cell density together with the probability of a cell dying p_{ij} and so will also be treated as a Poisson count. The probability of a cell dying is assumed to have a logistic relationship against dose with two parameters: slope B_i and location, X_i^* where X_i^* is the LC90.

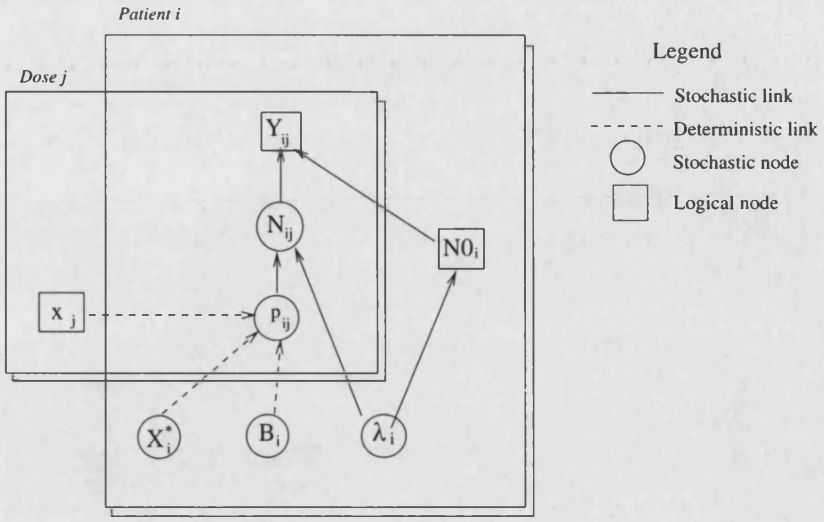

Figure 4-1: *Directed Acyclic Graph of initial model.*

4.2.1 Graphical model

A directed acyclic graph showing the conditional dependencies of the variables detailed above is shown in Figure 4-1. The variables are represented by *nodes.* If the variable contains actual data collected then it is designated as a *logical* or *deterministic* node. Conversely a *stochastic* node denotes a variable representing unobserved information: it is the posterior distribution of these variables that the analysis is designed to estimate. In the graph, square nodes represent actual data and the circular ones indicate stochastic variables. The levels of indexing are represented by the boxes: for each drug there are n patients, with results on 5 dose levels. At this stage the doses are assumed constant for each drug.

Purely deterministic or logical links are shown by dashed lines, stochastic dependencies by solid links.

4.2.2 The prior model

Since we are using a Bayesian analysis, each of the variables at the "top" of the model has to be described by a prior distribution. This distribution does not necessarily have to be a standard one but should describe the data adequately.

We would like the prior distributions to be sufficiently diffuse so that it gives reasonable weight to whichever region is supported by the data likelihood. Roberts and Smith (1994) give as a sufficient condition that these distributions coincide. Ideally the posterior distribution will be influenced little by the prior distribution and achieve convergence according to the data available.

Control value, $N0_i$

We assume that on the control slide, the cells are distributed according to a Poisson process with underlying density λ_i per unit area which is allowed to differ between patients. Since $N0_i$ is the average number of control cells per unit area over 10 such unit areas, the conditional distribution for $N0_i$ given λ_i is thus given as $1/10$ Pois$(10\lambda_i)$.

Underlying density of control cells, λ_i

From observation of previous assay results, the mean number of control cells is around 100 with a wide variation. Owing to such variation between patients, rather than have a common value, each patient has a separate λ. Since λ_i can take any positive real value, it is assigned a Gamma distribution $(\Gamma(2, \frac{1}{50}))$ which has mean 100 and variance of 5000 which is large enough to ensure sufficient probabilty over the range of likely values for N_0, for which the observed values range from < 1 to over 200.

Number of surviving cells, N_{ij}

The Y_{ij} depends upon the number of surviving cells so this must also be modelled and given a prior distribution. The numbers of surviving cells are not recorded but each counter will ascertain this value in order to calculate the proportion surviving after the various doses of drugs have been added. Cells may be counted accurately at low survival since interest is focused around the 90% kill rate. Unlike the $N0_i$, only one grid in the microscope eyepiece is used to estimate this value and so the distribution of surviving cells before the addition of any drug

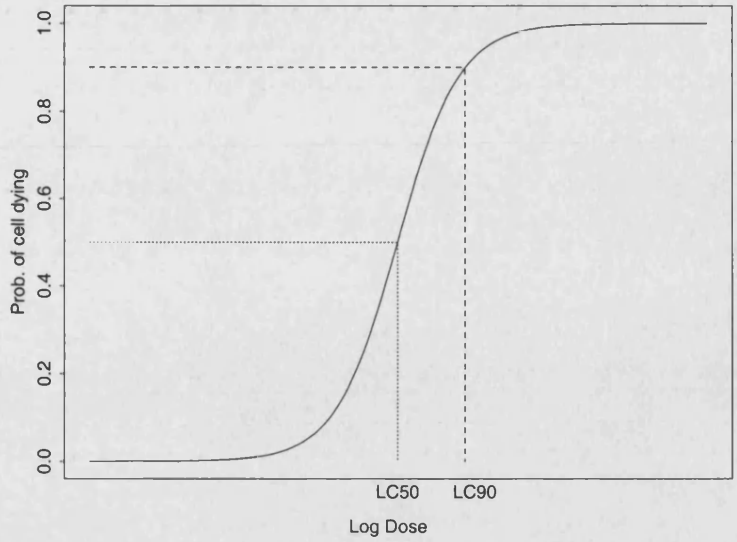

Figure 4-2: *Probability of a cell dying*

is given as Pois(λ_i). After drugs have been added, the N_{ij} now depend on two things: the expected number of control cells and the probability of a cell dying. The first of these is governed by λ_i and the second by p_{ij}. Thus N_{ij} is given a conditional distribution given λ_i and p_{ij} of Pois($(1 - p_{ij})\lambda_i$).

Probability of a cell dying, p_{ij}

The probability of a cell dying p_{ij} at a particular dose, x_{ij} , is modelled logistically since when no drug is added, it is assumed that all cells survive and at a sufficiently high dose all cells are killed. A typical example cell morbidity is shown in Figure 4-2.

There are two parameters which control the shape of this curve, the location and slope. We suppose the probability p of a cell dying at $\log_{10}$ dose x satisfies

$$\log \frac{p}{1-p} = a + bx, \tag{4.1}$$

We can re-express this as:

$$\log \frac{p}{1-p} = a + bx^* + b(x - x^*). \tag{4.2}$$

where x^* is the LC90 for the particular patient. At $x = x^*$ we have $p = 0.9$ so we now have

$$\log \frac{0.9}{0.1} = a + bx^*. \tag{4.3}$$

Substituting this back into Equation 4.2 we obtain

$$\log \frac{p_{ij}}{1-p_{ij}} = \log 9 + b(x_i - x_i^*), \tag{4.4}$$

which is a more convenient form to work with since we are primarily interested in estimating the LC90.

Since p_{ij} is a purely arithmetic function calculated for convenience as a parameter for the prior distribution of N_{ij}, it does not require any conditional distribution to be specified.

Slope of the cell survival curve, B_i

The slope of the cell survival curve indicates whether there is a sharp response or otherwise to the drug being tested over the assay range. Since there is little information about B_i and it is a "founder" node, it has been given a prior with a well dispersed Gamma distribution $\Gamma(\frac{1}{4}, \frac{1}{8})$ (mean=2, variance=16) to ensure it remains positive.

The dose, x_j

The dose is a predetermined value expressed on a $\log_{10}$ scale, calculated from the maximum concentration and dilution factors so that

$$x_j = \log_{10} \frac{c}{d^{(j-1)}} \quad \text{where } j = 1, ..., 5. \tag{4.5}$$

Since this was initially assumed to be constant for each drug, the doses were labelled $1, 2, ..., 5$, with "1" being the lowest dose.

The LC90. X_i^*

This is the value in which we are most interested since it indicates whether the drug being tested is likely to be effective for the patient. The units used depend upon the scale in use for the dose, x_j, (see below). This was given a fairly diffuse Normal prior distribution about the middle dose.

We define the prior distribution for X_i^* to be

$$N\left(\log_{10}(cd) - 3\log_{10}(d), \{\log_{10}(d)\}^2\right). \tag{4.6}$$

Since Equation 4.5 can be written as

$$x_j = \log_{10}(cd) - j\log_{10}(d) \quad \text{for } j = 1, ..., 5, \tag{4.7}$$

we see that this distribution is centred on x_3 with 2 standard deviations giving x_1 and x_5. This seems a reasonable model since the range of dose levels was selected to include the LC90 for most subjects.

Category for proportion of surviving cells, Y_{ij}

The Y_{ij} are derived from the number of surviving cells divided by the number of cells on the control slide. The resulting estimated proportion of surviving cells $N_{ij}/N0_i$ at dose level j is categorised and represented by Y_{ij}. Hence, Y_{ij} can be described as a variable taking a categorical distribution depending on the proportion of surviving cells. The categories are not precise and do overlap; in order to account for this, the probability of Y_{ij} being recorded as a particular category depends on whether $N_{ij}/N0_i$ falls within one of these overlapping ranges. The schematic diagram in Figure 4-3 gives a grapical representation showing all categories. If the proportion of surviving cells falls outside the allowable range of a category completely the probability of assignment to that category is zero.

Variable	Prior \ Likelihood Distribution	Description
Y_{ij}	$\text{Cat}(n_{ij}/n0_i)$	Recorded category of proportion surviving cells
$N0_i$	$\text{Pois}(\lambda_i)$	No. of control cells
N_{ij}	$\text{Pois}((1-p_{ij}\lambda_i)$	No. surviving cells at dose j
p_{ij}	*Deterministic*	Probability of a cell dying at dose j
λ_i	$\Gamma(2, \frac{1}{50})$	Underlying rate for control cells
x_j	*Deterministic*	Dose
X_i^*	$N\left(X_3, \{\log_{10}(d)\}^2\right)$	LC90
B_i	$\Gamma(\frac{1}{4}, \frac{1}{8})$	Slope of cell survival curve

Table 4.2: *Variables included in the initial model with their prior or likelihood distributions.*

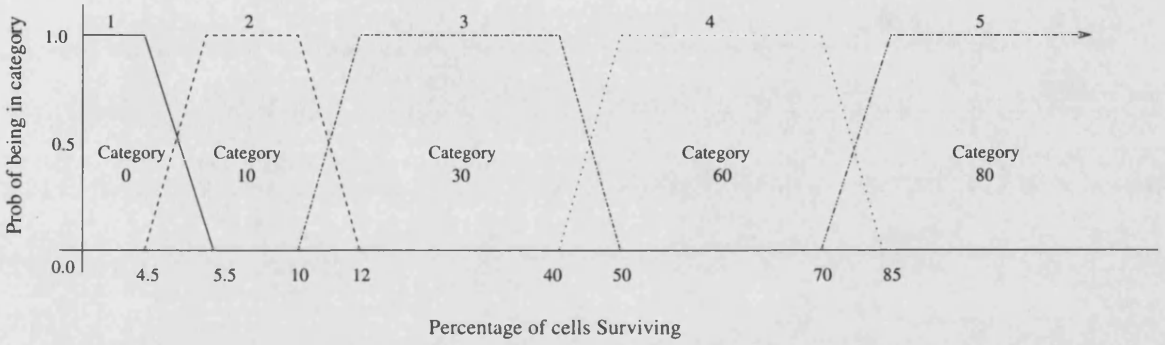

Figure 4-3: *Diagrammatic representation of probability of recorded category given the proportion of surviving cells (not to scale).*

Falling within any undisputed, non-overlapping area, yields a probability of 1 to assignment to this category. If the proportion falls within overlapping catergories, the probability of a category being recorded is a linear function dependent on the how close the proportion is to the category boundary.

For convenience, rather than use the initial category labels, the Y_{ij} take values of $1, ..., 5$ where 1 denotes the lowest category of 0-5.5% survival and 5 the highest at $> 70\%$.

To calculate the probability for, say, $P(Y_{ij} = 3 \mid N_{ij}/N0_i = r_{ij})$, the values are:

$$P(Y_{ij} = 3|\frac{N_{ij}}{N0_i} = r_{ij}) = \begin{cases} 0.0 & r_{ij} < 0.10 \\ \frac{r_{ij}-0.1}{0.12-0.1} & 0.10 \leq r_{ij} < 0.12 \\ 1.0 & 0.12 \leq r_{ij} < 0.40 \\ \frac{0.5-r_{ij}}{0.5-0.4} & 0.40 \leq r_{ij} < 0.50 \\ 0.0 & 0.50 \leq r_{ij} \end{cases}$$

The calculations are similar for all other values for Y_{ij} with the two extreme categories having just the 3 calculations. This distribution is denoted by the convention $Y_{ij} \sim Cat(n_{ij}/n0_i)$.

4.3 Model development

This initial model was used primarily as a test in order to develop the final model detailed below in Section 4.4. A random data set was generated and used to check whether the model as specified gave reliable results. Other test data from real examples was used to compare results with the manual graphical methods previously employed (See Section 2.2.3).

As more detailed information was received it became necessary to extend the model to incorporate the further categories for the proportion of cells surviving and the inclusion of more than one set of results for each assay where available.

It is also desirable for the prior distribution to be sufficiently dispersed so that it will allocate at least a moderate level of probability to the region in which the data likelihood is highest. If this is not the case, the posterior will be generated more by the prior than by the data and this is inappropriate in an application, such as this, where we have only weak prior beliefs. Inspection of both the prior and posterior distributions and the corresponding likelihood indicate that this is achieved with varying degrees of success. Obviously, success will vary across patients, disease and drug type. Figure 4-4 shows an example prior and posterior

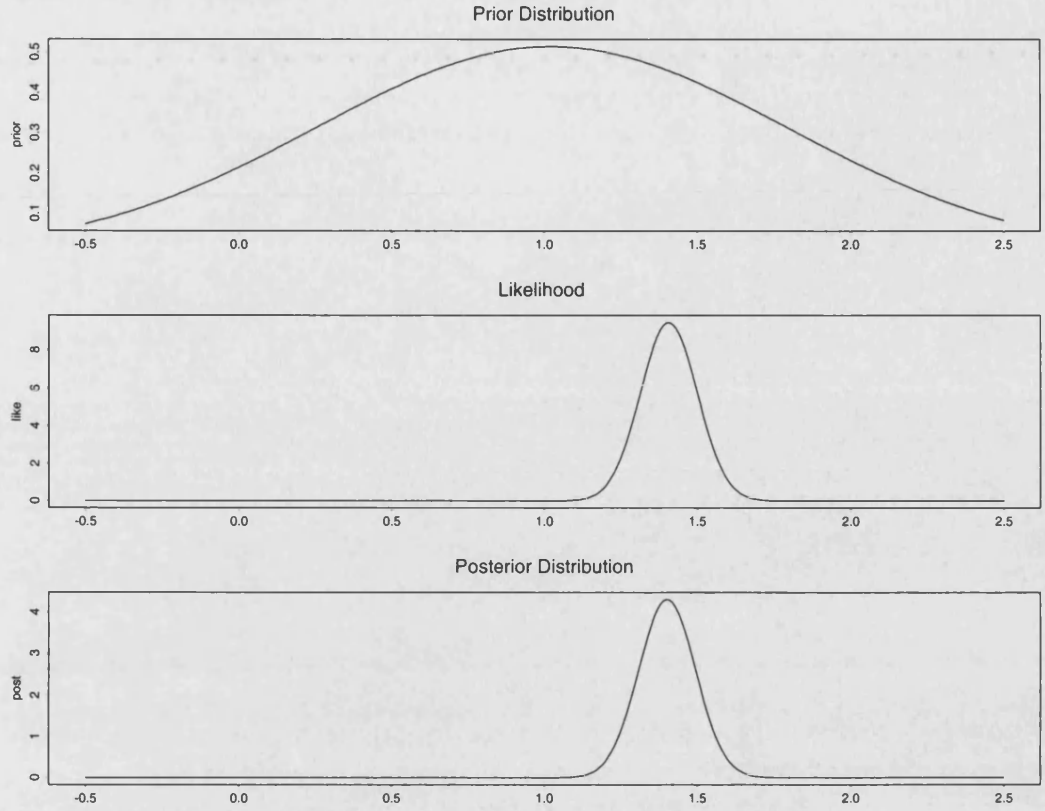

Figure 4-4: *Density plots for (i) the prior distribution (top), (ii) the likelihood (middle) and (iii) the resultant posterior distribution (bottom). Drug = TAX, variable =* $X^*(\log_{10} LC90)$*, patient=24.*

distribution together with the likelihood of the data (calculated up to a constant of proportionality).

In order to include information learned about founder nodes from the whole data set during the MCMC simulation, a further layer of nodes is added. Thus, for example, the slopes B_i may be modelled by a fairly compact distribution when there is evidence in the data that different subjects all have similar slopes, but the location for this compact distribution is given considerable freedom in the prior. As the analysis is conducted, information about slopes B_i for individuals is fed back through the model structure. This enables these nodes to depend even less on the prior distributions than before. These extra nodes are described in Section 4.4.4.

	Category name	Proportion of cells surviving
1)	0	0 - 2%
2)	5	0.1 - 8%
3)	10	2 - 12%
4)	15	10 - 25%
5)	30	15 - 45%
6)	45	30 - 60%
7)	60	45 - 100%
8)	100	> 60%

Table 4.3: *Categories used to denote proportion of surviving cells*

4.4 Development of final model

As further information and additional data were made available, it became necessary to make important changes to the initial basic model and the model in final use is described in this section.

Whilst the fundamental structure remains similar, the inclusion of *all* recorded results rendered necessary a further level of indexing. There are now patients $i = 1, ..., n$, doses $j = 1, ..., 5$ and replicates $k = 1, ...4$. Extra nodes were also added as the data structure gained complexity.

4.4.1 Additional cell survival categories

In order to record data more accurately, more categories indicating cell survival have been introduced. In total eight categories were used, with greater attention still being paid to low cell survival rates. The new categories are given in Table 4.3 and depicted in Figure 4-5. The probability calculations for each of these values with the additional categories are shown in Table 4.4.

$r_{ijk} = N0_{ik}/N_{ijk}$	$Pr(Y_{ijk} = y_{ijk} \mid r_{ijk} = N0_{ik}/N_{ijk})$	
0.0 – 0.001	$P(Y_{ijk} = 1) = 1.0$	
0.001 – 0.05	$P(Y_{ijk} = 1) = \frac{0.05 - r_{ijk}}{0.05 - 0.001}$	$P(Y_{ijk} = 2) = \frac{r_{ijk} - 0.001}{0.05 - 0.001}$
0.05 – 0.08	$P(Y_{ijk} = 2) = \frac{0.08 - r_{ijk}}{0.08 - 0.05}$	$P(Y_{ijk} = 3) = \frac{r_{ijk} - 0.05}{0.08 - 0.05}$
0.08 – 0.10	$P(Y_{ijk} = 3) = 1.0$	
0.10 – 0.12	$P(Y_{ijk} = 3) = \frac{0.12 - r_{ijk}}{0.12 - 0.10}$	$P(Y_{ijk} = 4) = \frac{r_{ijk} - 0.10}{0.12 - 0.10}$
0.12 – 0.15	$P(Y_{ijk} = 4) = 1.00$	
0.15 – 0.25	$P(Y_{ijk} = 4) = \frac{0.25 - r_{ijk}}{0.25 - 0.15}$	$P(Y_{ijk} = 5) = \frac{r_{ijk} - 0.15}{0.25 - 0.15}$
0.25 – 0.30	$P(Y_{ijk} = 5) = 1.0$	
0.30 – 0.45	$P(Y_{ijk} = 5) = \frac{0.45 - r_{ijk}}{0.45 - 0.30}$	$P(Y_{ijk} = 6) = \frac{r_{ijk} - 0.30}{0.45 - 0.30}$
0.45 – 0.60	$P(Y_{ijk} = 6) = \frac{0.60 - r_{ijk}}{0.60 - 0.45}$	$P(Y_{ijk} = 7) = \frac{r_{ijk} - 0.45}{0.60 - 0.45}$
0.60 – 1.00	$P(Y_{ijk} = 7) = \frac{1.00 - r_{ijk}}{1.00 - 0.60}$	$P(Y_{ijk} = 8) = \frac{r_{ijk} - 0.60}{1.00 - 0.60}$
> 1.00	$P(Y_{ijk} = 8) = 1.0$	

Table 4.4: *Probability calculations for* $Pr(Y_{ijk} = y_{ijk} \mid r_{ijk} = N_{ijk}/N0_{ik})$. *Categories not shown have zero probability.*

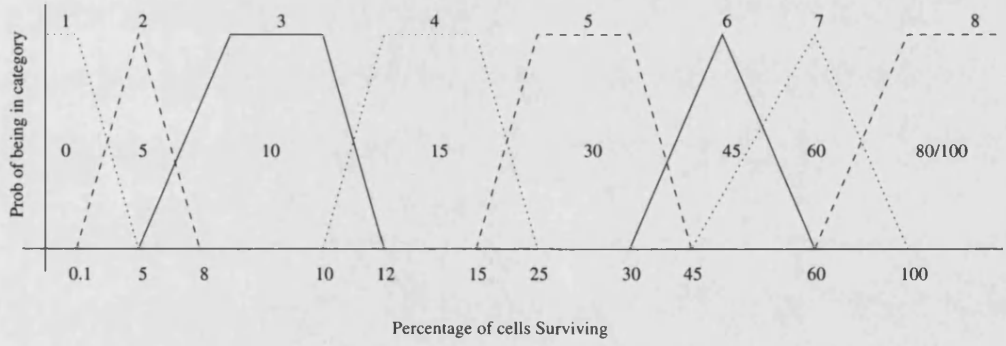

Figure 4-5: *Diagrammatic representation of the eight categories of surviving cells and their probabilities given* $n_{ijk}/n0_{ik}$ *(not to scale).*

Replicate	Slide	Counter
1	1	1
2	1	2
3	2	1
4	2	2

Table 4.5: *Table showing ordering of replicates, slides and counters.*

4.4.2 Replication of assay results

The new model incorporates information from all the data collected and not just one set of assay results. From Chapter 2, it is stated that up to four sets of results are recorded from the assays performed. Since there are two assays performed and two scientists independently categorising cell survival, a further level of indexing is required for each patient. The first two replicates concern slide one counted twice by different scientists. The third and fourth replicates concern the second slide again counted twice with the same counters as for slide one. See Table 4.5 for the ordering of indices.

There are now up to two control values supplied which have to be to dealt with. While initially only one value was given, in the later data, up to two values are recorded, one for each assay performed. If the second value is similar to the first then only the initial value is given, otherwise there are two values. Each scientist works with his or her own control values but that of the second scientist is not recorded. Hence $N0_{ik}$, $k = 1, .., 4$, can now have up to 4 values per patient depending on how many sets of assay results are noted, but only up to 2 control values are recorded for any one patient: $N0_{i1}$ is always given. Where replicate 3 exists, $N0_{i3}$ is either given or assigned to be the same value as $N0_{i1}$ since if the two control values are similar, only the first one is noted. The values for $N0_{i2}$ and $N0_{i4}$ are not given and are therefore treated as stochastic nodes and therefore simulated in order to obtain estimates for them.

There are thus up to four sets of results for the control values $N0_{ik}$, the number of surviving cells, N_{ijk} and the proportion surviving Y_{ijk}. Since the control values can therefore differ between each replicate, the value for λ, the underlying cell density, was allowed to differ within patients as well as between them. Whereas in

the initial model, these values were independent between patients, they are now only conditionally independent given the relevant *hyper-prior*, see Section 4.4.4 below. The λ_{ik}s show good agreement within patient but, as in the control value, there is considerable variation across *all* patients.

4.4.3 Variable dose levels

The five doses of a particular drug were initially held to be constant over all subjects, however, dose levels have changed as further knowledge about their behaviour became available. The nodes x_{ij}, $j = 1, ..., 5$ are still founder data nodes but their values are now allowed to differ between patients.

4.4.4 Hyper-priors

The final model thus has essentially the same structure as the initial model but with a further level of indexing required to incorporate all the replicates. Extra nodes have been added at the founder level in order to obtain a better model. These nodes are denoted *hyper-priors* and correspond to the parameters required to describe the prior distribution of the initial founder nodes.

We would like the prior distributions for the original founder nodes not to be too influential in the final summaries: it is useful to incorporate information gleaned from the data in the form of fairly informative priors rather than use improper prior distributions. In this way, by letting the distributions of these nodes have a stochastic dependency on the actual parameters of their prior distributions, there can be a "feedback" of information which is not possible in the initial model.

For the founder nodes with a Gamma prior distribution, λ_{ij} and B_i, the shape parameter was fixed and the scale parameters, θ_λ and θ_b, respectively were allowed to vary in order to give a fairly diffuse conditional prior distribution. The other founder node, X_i^* had been given a Normal distribution and hence has two hyper-prior parameters associated with it: μ_{x^*} and $\sigma^2_{x^*}$.

The new hyper-prior nodes themselves require prior distributions and these are

Variable	Prior Distribution	Description
Y_{ijk}	$\text{Cat}(n_{ijk}/n0_ik)$	Recorded category of proportion surviving cells
$N0_{ik}$	$\frac{1}{10}M_{ik}$	Average no. of control cells per unit area
M_{ik}	$\text{Pois}(10\lambda_i)$	No. of control cells in 10 unit areas
N_{ijk}	$\text{Pois}((1-p_{ijk})\lambda_i)$	No. surviving cells at dose j in 1 unit area
p_{ij}	*Deterministic*	Prob. of a cell dying at dose j function of X_i^*, B_i, x_{ij}
λ_i	$\Gamma(2, \frac{1}{\theta_\lambda})$	Underlying density of control cells
θ_λ	$\Gamma(4, 0.8)$	Hyper-prior: scale parameter for distn of λ_i
x_{ij}	*Deterministic*	Dose
X_i^*	$N(\mu_{x^*}, \sigma_{x^*}^2)$	LC90
μ_{x^*}	$N(0, 100)$	Hyper-prior: mean for X_i^*
$\sigma_{x^*}^{-2}$	$\Gamma(0.01, 0.01)$	Hyper-prior: variance for X_i^*
B_i	$\Gamma(\frac{1}{4}, \frac{1}{\theta_b})$	Slope of cell survival curve
θ_b	$\Gamma(1, \frac{1}{2})$	Hyper-prior: scale parameter for distn of b_i

Table 4.6: *Variables included in the final model with their prior distributions.*

given in Table 4.6. The distribution for μ_{x^*} has been chosen as a N(0,100), whilst the other nodes are constrained to remain positive and have been given Gamma distributions (see Table 4.6 for parameter values).

The conditional dependencies do not alter except at the previous founder level. These nodes are no longer founders since they are now conditional on their respective parents: the hyper-priors which have now become the new founder nodes.

The altered graphical model showing all the variables now included is shown in Figure 4-6.

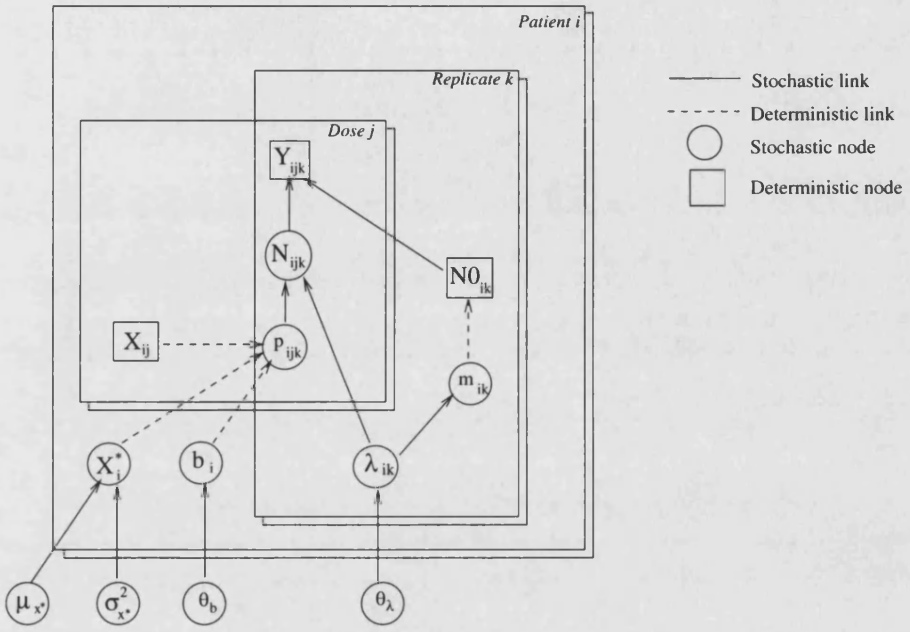

Figure 4-6: *Directed Acyclic Graph of final model.*

Chapter 5

Model Implementation

Having now defined a model which describes the data, the next task is to implement it. There are various options available and two approaches were actually used. The first approach was to use a publicly available package and the second meant writing a customised program. Obviously writing a program specifically designed for this problem meant being able to incorporate all the features of the data required. Despite the extra time required to develop such software, the gains in terms of flexibility to add extra features and make changes easily make this the more useful option.

5.1 BUGS

The initial model as defined in Chapter 4 was specified in the computer package *BUGS* - Bayesian inference Using Gibbs Sampling (Spiegelhalter *et al.* 1994). This package assumes a full probability model in which all quantities are treated as random variables. The data are then conditioned on in order to obtain a posterior distribution. Marginalising over this distribution in order to obtain inferences on the main quantities of interest is carried out using Gibbs sampling.

This program allows you to calculate summaries and examine the resulting chains generated. However, care needs to be take as convergence is not guaranteed in any fixed number of iterations. Resultant chains need to be saved and stored

in order to be able to assess their reliability properly. Separate diagnostics need to be performed on these chains to ensure convergence and determine whether the chain has been mixing well. Simple graphs are available within the package but more sophisticated techniques need to be done externally. One such package capable of providing implementations of several convergence diagnostics is CODA (Best *et al.* 1996).

Although BUGS, and its successor `winBUGS`, appeared to offer an initial solution to our problem and was used to test the aptness of the initial model using a randomly generated test data set, it was too slow to be of use. Given the vast number of runs to be made, efficiency was a major concern which meant looking at alternative methods.

5.2 Customised program - initial version

The alternative to using a commercially available package is to write a problem specific program which is capable of dealing with all the features of the data and is easily extendable.

Since not all the prior distributions specified were conjugate, these full conditional distributions could not be evaluated conveniently and hence a Gibbs sampler was not used. The rather cumbersome non-standard prior distribution for Y_{ijk} made the Metropolis Hastings method a far more economical algorithm to implement.

This program was written in ANSI C and initially tested using a randomly generated data set. Once this was found to be working satisfactorily, the first available data set was then used. The program gave intuitive results, the estimated LC90s being calculated at a slightly lower dosage than the manual graphical technique previously employed as described in Chapter 2 where the assay results were plotted, a curve fitted through the ranges of cell survival, and the LC90 was estimated by reading from the x-axis the dose at which there was estimated 10% cell survival. MCMC methods enable us to calculate estimates of the posterior intervals.

When the data set was finalised, some of the initial assumptions had to be changed

and the model was extended as discussed in Section 4.4 in order to incorporate these changes. "Hyper-priors" were introduced — a further set of founder nodes at the top level of the model with very diffuse prior distributions which enabled the previous founder nodes to be data driven and to reduce the influence of the prior distributions.

The model for the data has undergone several developments from the initial one specified and it is on the final model, as shown in Figure 4-6, that the rest of the chapter is based.

5.3 Program Algorithm

The program breaks down into 4 main sections:

- Read in data into appropriate structures
- Initialise variables
- Updating the Markov chain
- Compute and report results

5.4 Reading data

Since each drug is dealt with separately, only those records relating to the current drug are stored. Data are read in to appropriate structures and indicator variables are used to denote missing data values which are then treated in the same manner as a stochastic node.

5.5 Initialisation

Each stochastic node and missing data value has to be initialised before any analysis can be attempted. This is done from the "bottom" level upwards.

Starting values for each of the founder nodes, i.e., θ_λ, θ_b, μ_{x^*} and $\sigma^2_{x^*}$, were generated randomly from their prior distributions using methods described by Ripley (1987).

The offspring to these nodes (λ_i, b_i, x_i^*) also have their initial values generated from their prior distributions but allowing the values from their *hyper-priors* to be utilised.

Initialisation of $N0_{ik}$ differs depending on which replicate is being considered and whether there are one or two control values given, see Section 4.4.2. Replicate 1 is always given: if there is a second replicate, then $n0_{i2}$ is initialised to have the same value as $n0_{i1}$. If the third replicate exists and $n0_{i3}$ is not given, then it is assumed to have the same value as $n0_{i1}$ otherwise a separate value would be recorded. The fourth replicate is initially given the same value as $n0_{i3}$ if it is available or $n0_{i1}$ otherwise.

The underlying number of control cells over 10 areas, m_{ik}, is initialised to be $10 \times n0_{ik}$.

The probability of a cell dying p_{ijk} is a straightforward calculation involving x_{ij}, X_i^* and B_i (see Equation 4.4).

Once the $n0_{ik}$s have been determined, together with the category for the proportion of surviving cells, y_{ijk}, the number of surviving cells, n_{ijk}, can then be given starting values. The n_{ijk}s have constraints governing their legitimate range given by the extreme values of the survival categories, y_{ijk}, multiplied by $n0_{ik}$. Starting values are thus set to the mid-point of this range. If the recorded category is "8", i.e. representing $\geq$ 60% of surviving cells, there is no upper bound, and so starting values in this case range from $0.6 \times n0_{ik} + 20$. Although the value "20" is somewhat arbitrary, subsequent analyses show this value is unimportant.

The survival categories, y_{ijk}, are mostly recorded. Missing observations are set to the value of next replicate with a non-missing value for that particular drug dose. If there are no such values then this is set to the category for the replicate at the previous dose level. We could remove such Y_{ijk} from the graphical model but it is convenient to treat these as random variables and update them as the

notation remains consistent and it facilitates the estimation of the cell survival curve and hence the LC90.

5.6 Updating in the Markov Chain

Once all the values have been initialised, the Markov chain can then be constructed. The Metropolis-Hastings Algorithm as described in Section 3.3.1 has been used.

5.6.1 Proposals

Proposal distributions depend on the type of node being updated. If jumps are too small, the simulation moves very slowly through the target distribution; conversely big jumps will be rarely accepted, slowing the rate of convergence.

For a continuous variable with no constraints, a Normal distribution, with mean 0 and variance selected to ensure acceptance rates averaging around 40% is used to generate a proposal value. Whilst Müller (1993) proposed a scheme of accepting proposals with rates between 20%–80%, Roberts (1995) gives a guideline of 15%–50% in preference to this. Gelman *et al.* (1996a) proceed to show an optimal acceptance rate is roughly 40% and give a theoretical justification in Gelman *et al.* (1996b) but concede other schemes work as well but may converge more slowly. In practice, precise acceptance rates are unattainable for every simulation being carried out and therefore, of the variables monitored, the target acceptance rate around 40% was not always achieved with some rates differing considerably. This was usually due to the restrictive range limits of legitimate values especially for discrete variables.

For variables with a Gamma prior distribution, a Normal distribution is still used to generate proposals but negative proposed values are discarded and a new proposal generated if necessary. This effectively gives a left truncated Normal distribution and care needs to be taken when calculating transition probabilities in this case. We used a Normal kernal here to produce proposal values since

generating a random variable with a gamma distribution was more expensive computationally especially when there is a non-integer shape parameter.

Proposal values for discrete variables are generated to be uniformly distributed in the range $x-k,...,x-1,x+1,...,x+k$, where x is the current variable and k is the largest sized jump allowed; this value is predetermined to give reasonable acceptance rates. If $x \pm k$ falls outside the allowable range, k is decreased so there will be two different values, k_{min} and k_{max}. Proposals, x' will then fall within $(x-k_{min}) \leq x' \leq (x+k_{max})$. The possible proposed values are still uniformly distributed either side of x but are no longer uniformly distributed over the whole range so that: $P((x-k_{min}) \leq X' < x) = 0.5$ and $P((x < X' \leq (x+k_{max})) = 0.5$.

Occasionally the range of valid values is so narrow so that proposals are necessarily the same as the current value, this may happen if both $N0_{ik}$ and Y_{ijk} have small values. In these cases, no update is attempted.

Having non-symmetric proposal distributions will affect the transition probability calculations (see below).

5.6.2 Acceptance probabilities

The acceptance probability, α, of a proposed value, x' say, is the product of two quantities: the ratio of the posteriors with regard to the proposed value and the ratio of the transition probabilities. If this quantity is greater than 1, α takes the value of 1, i.e.:

$$\alpha(x,x') = \min\left(1, \frac{\pi(x')q(x',x)}{\pi(x)q(x,x')}\right), \tag{5.1}$$

where $\pi(x')$ is the product of the prior and the likelihood of the data and $q(x',x)$ is the probability of proposing a value x from x'.

If the proposal distribution is symmetric, then the transition probabilities are equal and therefore do not need to be evaluated as they cancel. However, proposals with a non-symmetric distribution, i.e. those with range constraints,

will still need the transition probabilities to be calculated.

5.6.3 Transition probabilities

If the variable has a Gamma prior distribution then proposals need to be positive. A proposal is generated by adding to the current value, for example $\sigma_{x^*}^{-2}$, a randomly generated value from a Normal distribution with mean zero and standard deviation s to obtain $\sigma_{x^*}^{-2'}$. If $\sigma_{x^*}^{-2'}$ is negative another random number is generated until the proposed value is valid. Clearly these proposed values do not form a standard Normal distribution since this distribution is effectively truncated at its lower tail since negative values of a magnitude greater than $\sigma_{x^*}^{-2}$ are invalid. Therefore, in order to compute

$$q(\sigma_{x^*}^{-2'}, \sigma_{x^*}^{-2})$$

we need to re-scale the proposal distribution to have unit area since values are invalid over its lower tail.

$$\begin{aligned} q(\sigma_{x^*}^{-2}, \sigma_{x^*}^{-2'}) &= \frac{\frac{1}{\sqrt{s}}\phi\left(\frac{\sigma_{x^*}^{-2'} - \sigma_{x^*}^{-2}}{s}\right)}{\Phi\left(\frac{\sigma_{x^*}^{-2}}{s}\right)} \\ &= \frac{\frac{1}{\sqrt{2\pi s^2}} \exp\left[-\frac{(\sigma_{x^*}^{-2} - \sigma_{x^*}^{-2'})^2}{2s^2}\right]}{\Phi\left(\frac{\sigma_{x^*}^{-2}}{s}\right)}. \end{aligned}$$

This now gives us the following:

$$\frac{q(\sigma_{x^*}^{-2'}, \sigma_{x^*}^{-2})}{q(\sigma_{x^*}^{-2}, \sigma_{x^*}^{-2'})} = \frac{\left[\Phi\left(\frac{\sigma_{x^*}^{-2'}}{s}\right)\right]^{-1} \frac{1}{\sqrt{2\pi s^2}} \exp\left[-\frac{(\sigma_{x^*}^{-2'} - \sigma_{x^*}^{-2})^2}{2s^2}\right]}{\left[\Phi\left(\frac{\sigma_{x^*}^{-2}}{s}\right)\right]^{-1} \frac{1}{\sqrt{2\pi s^2}} \exp\left[-\frac{(\sigma_{x^*}^{-2} - \sigma_{x^*}^{-2'})^2}{2s^2}\right]}$$

$$= \frac{\Phi\left(\frac{\sigma_{x^*}^{-2}}{s}\right)}{\Phi\left(\frac{\sigma_{x^*}^{-2'}}{s}\right)}.$$

The cdf of the standardised Normal distribution is computed using the algorithm given by Hastings (1955) based on a polynomial approximation $P(z) = \Phi(z) + \epsilon$. The error term, ϵ, is such that $|\epsilon(x)| < 1.5 \times 10^{-7}$.

For discrete variables, a similar approach is adopted in that there are often range limits. The maximum proposal size is predetermined according to the valid range in order to give acceptance rates of around 40% (see Section 5.6.1). The proposed value, x' say, is obtained by adding a uniform random variable to x over the range $\{-k, ..., -1, 1, ..., +k\}$. If the maximum jump size allowed takes the proposed value outside the valid range then this probability calculation is modified over the range $\{k_{min}, ..., -1, 1, ..., +k_{max}\}$ so that $P(x' > x) = P(x' < x) = 0.5$. So, suppose, for example $x' > x$ where x' has $k_{max}(x')$ and x has $k_{min}(x)$, then

$$\begin{aligned} \frac{q(x', x)}{q(x, x')} &= \frac{0.5/k_{max}(x')}{0.5/k_{max}(x'))} \\ &= \frac{k_{min}(x)}{k_{max}(x')} \end{aligned}$$

If the current value is already at the minimum allowed value then the proposed value can obviously be only greater than this, and analogeously for the maximum.

5.6.4 Acceptance probability calculations

For computational ease, logs have been taken of the posterior part of the expression Equation 5.1 to reduce a product to a sum which can be computed more efficiently as computer addition is significantly quicker than multiplication (Press *et al.* 1988).

Notation

The following conventions have been used throughout the following section.

P – Number of patients

D – Number of doses - this is always 5

R_i – Number of replicates for patient i (up to 4)

$\pi(x)$ – Likelihood of the dataset and current parameter values

p_{ijk} – Probability of a cell dying for patient i, dose j, replicate k

$$= \frac{9 \exp\left[b_i(x_{ijk} - x_i^*)\right]}{10 + \exp\left[b_i(x_{ijk} - x_i^*)\right]}.$$

Where p'_{ijk} is used, this denotes a substitution of one of the variables b_i or x_i^* by its proposed value b' or x'^*.

5.6.5 Full conditional probability distribution

It is convenient to state the conditional probability for the full model given the observed data from which subsequent calculations are derived. The probability of a given set of observations can be written as:

$$\pi(\mu_{x^*}, \sigma^2_{x^*}, \theta_b, \theta_\lambda, x_i^*, b_i, x_{ij}, \lambda_{ik}, m_{ik}, n0_{ik}, p_{ijk}, n_{ijk}, y_{ijk})$$

$$\begin{aligned}
= \; & f_{\mu_{x^*}}(\mu_{x^*}) f_{\sigma^2_{x^*}}(\sigma^2_{x^*}) f_{\theta_b}(\theta_b) f_{\theta_\lambda}(\theta_\lambda) \\
& \prod_{i=1}^{P} \Big\{ f_{X_i^*|\sigma^2_{x^*},\mu_{x^*}}(x_i^*|\sigma^2_{x^*}, \mu_{x^*}) f_{B_i|\theta_b}(b_i|\theta_b) \\
& \prod_{k=1}^{R_i} \left[f_{\lambda_{ik}|\theta_\lambda}(\lambda_{ik}|\theta_\lambda) f_{M_{ik}|\lambda_{ik}}(m_{ik}|\lambda_{ik}) f_{N0_{ik}|M_{ik}}(n0_{ik}|m_{ik}) \right. \\
& \prod_{j=1}^{D} \left(f_{P_{ijk}|X_{ik},X_i^*,B_i}(p_{ijk}|x_{ik}, x_i^*, b_i) f_{N_{ijk}|P_{ijk},\lambda_{ik}}(n_{ijk}|p_{ijk}, \lambda_{ik}) \right. \\
& \left. \left. f_{Y_{ijk}|N_{ijk},N0_{ik}}(y_{ijk}|n_{ijk}, n0_{ik}) \right) \right] \Big\}. \qquad (5.2)
\end{aligned}$$

The conditional probability of unknown variables given observed data is proportional to this quantity. All further calculations are derived from this expression. Terms not involving the variable being updated cancel when forming ratios leaving a much reduced formula.

5.6.6 Acceptance probability for an update of $\sigma^2_{\mu^*}$

We wish to calculate the ratio of probabilities $\pi(x')/\pi(x)$ where the numerator refers to the probability of the whole dataset but substituting a proposal for the current value being updated, and the denominator refers to the probability of the current set of values.

Given $\sigma^{-2}_{x^*}$ has prior distribution $\sigma^{-2}_{x^*} \sim \Gamma(\tau, \upsilon)$, where τ and υ are constants with values: $\tau = 0.01, \upsilon = 0.01$, we get mean and variance of 1 and 100 respectively which gives a sufficient range for the parameter $\sigma^{-2}_{x^*}$. The ratio of likelihoods is computed as follows:

$$\begin{aligned}
\frac{\pi(x')}{\pi(x)} &= \frac{f_{\sigma^{-2}_{x^*}}(\sigma^{-2'}_{x^*}) \prod_{i=1}^{\mathrm{P}} f_{X^*_i|\sigma^2_{x^*},\mu_{x^*}}(x^*_i|\sigma^{2'}_{x^*}, \mu_{x^*})}{f_{\sigma^{-2}_{x^*}}(\sigma^{-2}_{x^*}) \prod_{i=1}^{\mathrm{P}} f_{X^*_i|\sigma^2_{x^*},\mu_{x^*}}(x^*_i|\sigma^2_{x^*}, \mu_{x^*})} \\
&= \frac{\frac{\upsilon^\tau}{\Gamma(\tau)} \sigma^{-2'(\tau-1)}_{x^*} \exp\left(-\upsilon\sigma^{-2'}_{x^*}\right) \prod_{i=1}^{\mathrm{P}} \frac{1}{\sqrt{2\pi\sigma^{2'}_{x^*}}} \exp\left(-\frac{1}{2}\frac{(x^*_i-\mu_{x^*})^2}{\sigma^{2'}_{x^*}}\right)}{\frac{\upsilon^\tau}{\Gamma(\tau)} \sigma^{-2(\tau-1)}_{x^*} \exp\left(-\upsilon\sigma^{-2}_{x^*}\right) \prod_{i=1}^{\mathrm{P}} \frac{1}{\sqrt{2\pi\sigma^2_{x^*}}} \exp\left(-\frac{1}{2}\frac{(x^*_i-\mu_{x^*})^2}{\sigma^2_{x^*}}\right)}
\end{aligned}$$

Taking logs:

$$\begin{aligned}
\log\frac{\pi(x')}{\pi(x)} &= (\tau-1)(\log\sigma^{-2'}_{x^*} - \log\sigma^{-2}_{x^*}) + \upsilon(\sigma^{-2}_{x^*} - \sigma^{-2'}_{x^*}) \\
&\quad + \sum_{i=1}^{\mathrm{P}} \left\{ \frac{1}{2}(\log\sigma^{-2}_{x^*} - \log\sigma^{-2'}_{x^*}) + \frac{1}{2}(x^*_i - \mu_{x^*})^2 \left(\frac{1}{\sigma^2_{x^*}} - \frac{1}{\sigma^{2'}_{x^*}}\right) \right\} \\
&= \left(\frac{\mathrm{P}}{2} + 1 - \tau\right) \left(\log\frac{\sigma^{2'}_{x^*}}{\sigma^2_{x^*}}\right) + \left(\upsilon + \frac{1}{2}\sum_{i=1}^{\mathrm{P}}(x^*_i - \mu_{x^*})^2\right) \left(\frac{1}{\sigma^2_{x^*}} - \frac{1}{\sigma^{2'}_{x^*}}\right)
\end{aligned}$$

We then multiply this by the transition probability ratio

$$\frac{q(x',x)}{q(x,x')} = \frac{\Phi\left(\frac{\sigma_{x^*}^{-2}}{s}\right)}{\Phi\left(\frac{\sigma_{x^*}^{-2'}}{s}\right)}$$

to obtain the acceptance probability of the proposed value. Here s is the standard deviation of the Normal kernel used to generate proposals.

The probability, $\alpha(x',x)$, of accepting the proposed value for $\sigma_{x^*}^{-2}$ is then:

$$\min\left(1, \left(\frac{\mathrm{P}}{2}+1-\tau\right)\left(\log\frac{\sigma_{x^*}^{2'}}{\sigma_{x^*}^2}\right) + \left(v+\frac{1}{2}\sum_{i=1}^{\mathrm{P}}(x_i^*-\mu_{x^*})^2\right)\left(\frac{1}{\sigma_{x^*}^2}-\frac{1}{\sigma_{x^*}^{2'}}\right)\frac{\Phi\left(\frac{\sigma_{x^*}^{-2}}{s}\right)}{\Phi\left(\frac{\sigma_{x^*}^{-2'}}{s}\right)}\right).$$

5.6.7 Acceptance probability for an update of μ_{x^*}

The prior distribution given to μ_{x^*} is $\mu_{x^*} \sim N(m,v)$. Hence the proposal values are not constrained and so have symmetric transition probabilities which cancel out and do not need to be calculated and so $\alpha = \min(1, \pi(x')/\pi(x))$.

$$\begin{aligned}
\frac{\pi(x')}{\pi(x)} &= \frac{f_{\mu_{x^*}}(\mu'_{x^*})\prod_{i=1}^{\mathrm{P}} f_{X_i^*|\sigma_{x^*}^{-2},\mu_{x^*}}(x_i^*|\sigma_{x^*}^{-2},\mu'_{x^*})}{f_{\mu_{x^*}}(\mu_{x^*})\prod_{i=1}^{\mathrm{P}} f_{X_i^*|\sigma_{x^*}^{-2},\mu_{x^*}}(x_i^*|\sigma_{x^*}^{-2},\mu_{x^*})} \\
&= \frac{\frac{1}{\sqrt{2\pi v}}\exp\left(-\frac{1}{2}\frac{(\mu'_{x^*}-m)^2}{v}\right)\prod_{i=1}^{\mathrm{P}}\frac{1}{\sqrt{2\pi v}}\exp\left(-\frac{1}{2}\frac{(x_i^*-\mu'_{x^*})^2}{\sigma_{x^*}^2}\right)}{\frac{1}{\sqrt{2\pi v}}\exp\left(-\frac{1}{2}\frac{(\mu_{x^*}-m)^2}{v}\right)\prod_{i=1}^{\mathrm{P}}\frac{1}{\sqrt{2\pi v}}\exp\left(-\frac{1}{2}\frac{(x_i^*-\mu_{x^*})^2}{\sigma_{x^*}^2}\right)}
\end{aligned}$$

Taking logs:

$$\begin{aligned}
\log\frac{\pi(x')}{\pi(x)} &= -\frac{1}{2v}(\mu'_{x^*}-m)^2 + \frac{1}{2v}(\mu_{x^*}-m)^2 \\
&\quad + \sum_{i=1}^{\mathrm{P}}\frac{-1}{2\sigma_{x^*}^2}(x_i^*-\mu'_{x^*})^2 + \frac{1}{2\sigma_{x^*}^2}(x_i^*-\mu_{x^*})^2
\end{aligned}$$

$$= \frac{1}{2v}\left[(\mu_{x^*} - m)^2 - (\mu'_{x^*} - m)^2\right] + \frac{1}{2\sigma^2_{x^*}}\sum_{i=1}^{P}\left[(x_i^* - \mu_{x^*})^2 - (x_i^* - \mu'_{x^*})^2\right]$$

The probability of accepting a proposed value for μ_{x^*} is thus:

$$\alpha = \min\left(1, \frac{1}{2v}\left[(\mu_{x^*} - m)^2 - (\mu'_{x^*} - m)^2\right] + \frac{1}{2\sigma^2_{x^*}}\sum_{i=1}^{P}\left[(x_i^* - \mu_{x^*})^2 - (x_i^* - \mu'_{x^*})^2\right]\right).$$

5.6.8 Acceptance probability for an update of θ_b

θ_b has been given prior distribution $\theta_b \sim \Gamma(\tau, v)$ with $\tau = \frac{1}{4}$ and $v = \frac{1}{40}$ giving a mean of 10 and a variance of 400 in order to ensure a sufficient range of suitable values for this parameter.

$$\frac{\pi(x')}{\pi(x)} = \frac{f_{\theta_b}(\theta'_b)\prod_{i=1}^{P} f_{B_i|\theta_b}(b_i|\theta'_b)}{f_{\theta_b}(\theta_b)\prod_{i=1}^{P} f_{B_i|\theta_b}(b_i|\theta_b)}$$

$$= \frac{\frac{v^\tau}{\Gamma(\tau)}\theta_b'^{(\tau-1)}\exp(-v\theta'_b)\prod_{i=1}^{P}\frac{1}{\theta_b'^k\Gamma(k)}b_i^{(k-1)}\exp\left(\frac{-b_i}{\theta'_b}\right)}{\frac{v^\tau}{\Gamma(\tau)}\theta_b^{(\tau-1)}\exp(-v\theta_b)\prod_{i=1}^{P}\frac{1}{\theta_b^k\Gamma(k)}b_i^{(k-1)}\exp\left(\frac{-b_i}{\theta_b}\right)}$$

Taking logs:

$$\log\frac{\pi(x')}{\pi(x)} = (\tau-1)\log\theta'_b - (\tau-1)\log\theta_b - v\theta'_b + v\theta_b + \sum_{i=1}^{P}\left(-k\log\theta'_b + k\log\theta_b - \frac{b_i}{\theta'_b} + \frac{b_i}{\theta_b}\right)$$

$$= (\tau - 1 - kP)\left(\log\frac{\theta'_b}{\theta_b}\right) + v(\theta_b - \theta'_b) + \left(\frac{1}{\theta_b} - \frac{1}{\theta'_b}\right)\sum_{i=1}^{P} b_i$$

The ratio of transition probabilities is:

$$\frac{q(\theta_b', \theta_b)}{q(\theta_b, \theta_b')} = \frac{\Phi\left(\frac{\theta_b}{s}\right)}{\Phi\left(\frac{\theta_b'}{s}\right)}$$

Where s is the standard deviation of the Normal kernel used to generate proposals. The acceptance probability α is calculated as before.

5.6.9 Acceptance probability for updating θ_λ

The prior distribution given to θ_λ is $\theta_\lambda \sim \Gamma(\tau, \upsilon)$

$$\begin{aligned}
\frac{\pi(x')}{\pi(x)} &= \frac{f_{\theta_\lambda}(\theta_\lambda') \prod_{i=1}^{P} \prod_{k=1}^{R_i} f_{\lambda_{ik}|\theta_\lambda}(\lambda_{ik}|\theta_\lambda')}{f_{\theta_\lambda}(\theta_\lambda) \prod_{i=1}^{P} \prod_{k=1}^{R_i} f_{\lambda_{ik}|\theta_\lambda}(\lambda_{ik}|\theta_\lambda)} \\
&= \frac{\frac{\upsilon^\tau}{\Gamma(\tau)} \theta_\lambda'^{(\tau-1)} \exp\left(-\upsilon\theta_\lambda'\right) \prod_{i=1}^{P} \prod_{k=1}^{R_i} \frac{1}{\theta_\lambda'^{\kappa_\lambda} \Gamma(\kappa_\lambda)} \lambda_{ik}^{(\kappa_\lambda - 1)} \exp\left(\frac{-\lambda_{ik}}{\theta_\lambda'}\right)}{\frac{\upsilon^\tau}{\Gamma(\tau)} \theta_\lambda^{(\tau-1)} \exp\left(-\upsilon\theta_\lambda\right) \prod_{i=1}^{P} \prod_{k=1}^{R_i} \frac{1}{\theta_\lambda^{\kappa_\lambda} \Gamma(\kappa_\lambda)} \lambda_{ik}^{(\kappa_\lambda - 1)} \exp\left(\frac{-\lambda_{ik}}{\theta_\lambda}\right)}
\end{aligned}$$

Taking logs:

$$\begin{aligned}
\log \frac{\pi(x')}{\pi(x)} &= (\tau - 1) \log \theta_\lambda' - (\tau - 1) \log \theta_\lambda - \upsilon\theta_\lambda' + \upsilon\theta_\lambda' \\
&\quad + \sum_{i=1}^{P} \sum_{j=1}^{R_i} \left(-\kappa_\lambda \log \theta_\lambda' + \kappa_\lambda \log \theta_\lambda - \frac{\lambda_{ik}}{\theta_\lambda'} + \frac{\lambda_{ik}}{\theta_\lambda}\right) \\
&= (\tau - 1 - PR\kappa_\lambda) \left(\log \frac{\theta_\lambda'}{\theta_\lambda}\right) + \upsilon(\theta_\lambda - \theta_\lambda') + \sum_{i=1}^{P} \sum_{j=1}^{R_i} \lambda_{ik} \left(\frac{1}{\theta_\lambda} - \frac{1}{\theta_\lambda'}\right)
\end{aligned}$$

The ratio of transition probabilities is:

$$\frac{q(\theta'_\lambda, \theta_\lambda)}{q(\theta_\lambda, \theta'_\lambda)} = \frac{\Phi\left(\frac{\theta_\lambda}{s}\right)}{\Phi\left(\frac{\theta'_\lambda}{s}\right)}$$

Where s is the standard deviation of the Normal kernel used to generate proposals. The acceptance probability α is calculated as before.

5.6.10 Acceptance probability for updating b_i

The prior conditional distribution for b_i is $\sim \Gamma(\kappa_b, \frac{1}{\theta_b})$ where the shape parameter $\kappa_b = 0.25$. The skew shape of this distribution makes it not improbable for the presence of steep slopes.

$$\begin{aligned}
\frac{\pi(x')}{\pi(x)} &= \frac{f_{B_i}(b'_i)\prod_{j=1}^{D}\prod_{k=1}^{R_i} f_{N_{ijk}|\lambda_{ik},B_i,X_i^*,X_{ij}}(n_{ijk}|\lambda_{ik}, b'_i, x_i^*, x_{ij})}{f_{B_i}(b_i)\prod_{j=1}^{D}\prod_{k=1}^{R_i} f_{N_{ijk}|\lambda_{ik},B_i,X_i^*,X_{ij}}(n_{ijk}|\lambda_{ik}, b_i, x_i^*, x_{ij})} \\
&= \frac{\frac{1}{\theta_b^\kappa \Gamma(\kappa_b)} b_i'^{(\kappa_b-1)} \exp\left(\frac{-b'_i}{\theta_b}\right)}{\frac{1}{\theta_b^\kappa \Gamma(\kappa_b)} b_i^{(\kappa_b-1)} \exp\left(\frac{-b_i}{\theta_b}\right)} \\
&\quad \times \frac{\prod_{j=1}^{D}\prod_{k=1}^{R_i} \left[\lambda_{ik}(1-p'_{ijk})\right]^{n_{ijk}} \exp\left[-\lambda_{ik}(1-p'_{ijk})\right] / n_{ijk}!}{\prod_{j=1}^{D}\prod_{k=1}^{R_i} \left[\lambda_{ik}(1-p_{ijk})\right]^{n_{ijk}} \exp\left[-\lambda_{ik}(1-p_{ijk})\right] / n_{ijk}!}
\end{aligned}$$

Taking logs :

$$\begin{aligned}
\log\frac{\pi(x')}{\pi(x)} &= (\kappa_b - 1)(\log b'_i - \log b_i) + \frac{1}{\theta_b}(b_i - b'_i) \\
&\quad + \sum_{j=1}^{D}\sum_{k=1}^{R_i} \left\{ n_{ijk}\left[\log(\lambda_{ik}(1-p'_{ijk})) - \log(\lambda_{ik}(1-p_{ijk}))\right] \right. \\
&\quad \left. + \lambda_{ik}\left[(1-p_{ijk}) - (1-p'_{ijk})\right] \right\}
\end{aligned}$$

$$= (\kappa_b - 1)\log\left(\frac{b_i'}{b_i}\right) + \frac{1}{\theta_b}(b_i - b_i') + \sum_{j=1}^{D}\sum_{k=1}^{R_i}\left\{n_{ijk}\log\left(\frac{1-p_{ijk}'}{1-p_{ijk}}\right) + \lambda_{ik}(p_{ijk}' - p_{ijk})\right\}$$

The ratio of transition probabilities is:

$$\frac{q(b_i', b_i)}{q(b_i, b_i')} = \frac{\Phi\left(\frac{b}{s}\right)}{\Phi\left(\frac{b'}{s}\right)}$$

Where s is the standard deviation of the Normal kernel used to generate proposals. The acceptance probability α is calculated as before.

5.6.11 Acceptance probability for updating X^*

The prior conditional distribution for X_i^* is $N(\mu_{x^*}, \sigma_{x^*}^2)$

$$\frac{\pi(x')}{\pi(x)} = \frac{f_{X^*|\mu_x^*,\sigma_{x^*}^2}(x_i^{*\prime}|\mu_{x^*},\sigma_{x^*}^2)\prod_{j=1}^{D}\prod_{k=1}^{R_i} f_{N_{ijk}|X_{ik},X_i^*,B_i,\lambda_{ik}}(n_{ijk}|x_{ik},x_i^{*\prime},b_i,\lambda_{ik})}{f_{X^*|\mu_x^*,\sigma_{x^*}^2}(x_i^*|\mu_{x^*},\sigma_{x^*}^2)\prod_{j=1}^{D}\prod_{k=1}^{R_i} f_{N_{ijk}|X_{ik},X_i^*,B_i,\lambda_{ik}}(n_{ijk}|x_{ik},x_i^*,b_i,\lambda_{ik})}$$

$$= \frac{\frac{1}{\sqrt{2\pi\sigma_{x^*}^2}}\exp\left\{-\frac{1}{2}\frac{(x_i^{*\prime}-\mu_{x^*})^2}{\sigma_{x^*}^2}\right\}}{\frac{1}{\sqrt{2\pi\sigma_{x^*}^2}}\exp\left\{-\frac{1}{2}\frac{(x_i^{*}-\mu_{x^*})^2}{\sigma_{x^*}^2}\right\}} \times \frac{\prod_{j=1}^{D}\prod_{k=1}^{R_i}\frac{[\lambda_{ik}(1-p_{ijk}')]^{n_{ijk}}}{n_{ijk}!}\exp\left\{-\lambda_{ik}(1-p_{ijk}')\right\}}{\prod_{j=1}^{D}\prod_{k=1}^{R_i}\frac{[\lambda_{ik}(1-p_{ijk})]^{n_{ijk}}}{n_{ijk}!}\exp\left\{-\lambda_{ik}(1-p_{ijk})\right\}}$$

Taking logs

$$\log\frac{\pi(x')}{\pi(x)} = -\frac{1}{2\sigma_{x^*}^2}(x_i^{*\prime} - \mu_{x^*})^2 + \frac{1}{2\sigma_{x^*}^2}(x_i^* - \mu_{x^*})^2$$

$$+\sum_{j=1}^{D}\sum_{k=1}^{R}\{n_{ijk}\log[\lambda_{ik}(1-p'_{ijk})]-n_{ijk}\log[\lambda_{ik}(1-p_{ijk})]$$
$$-\lambda_{ik}(1-p'_{ijk})+\lambda_{ik}(1-p_{ijk})\}$$

$$= \frac{1}{2\sigma^2_{x^*}}\left[(x_i^*-\mu_{x^*})^2-(x_i^{*\prime}-\mu_{x^*})^2\right]$$
$$+\sum_{j=1}^{D}\sum_{k=1}^{R}n_{ijk}\log\left[\frac{1-p'_{ijk}}{1-p_{ijk}}\right]+\lambda_{ik}(p'_{ijk}-p_{ijk})$$

Since X_i^* can take any value, the transition probabilities are symmetric and therefore do not need to be calculated as $q(x^{*\prime},x^*)/q(x^*,x^{*\prime})=1$.

5.6.12 Acceptance probability for updating λ_{ik}

The prior conditional distribution for λ_{ik}, the underlying rate of all cells, is $\lambda_{ik}\sim\Gamma(\kappa_\lambda,\frac{1}{\theta_\lambda})$.

$$\frac{\pi(x')}{\pi(x)} = \frac{f_{\lambda|\theta_\lambda}(\lambda'_{ik}|\theta_\lambda)\ f_{M|\lambda}(m_{ik}|\lambda'_{ik})\prod_{j=1}^{D}f_{N|\lambda,B,X^*,X}(n_{ijk}|\lambda'_{ik},b_i,x_i^*,x_{ij})}{f_{\lambda|\theta_\lambda}(\lambda_{ik}|\theta_\lambda)\ f_{M|\lambda}(m_{ik}|\lambda_{ik})\prod_{j=1}^{D}f_{N|\lambda,B,X^*,X}(n_{ijk}|\lambda_{ik},b_i,x_i^*,x_{ij})}$$

$$= \frac{\frac{1}{\theta_\lambda^{(\kappa_\lambda-1)}\Gamma(\kappa_\lambda)}\exp\left(\frac{-\lambda'_{ik}}{\theta_\lambda}\right)\lambda_{ik}^{\prime(\kappa_\lambda-1)}(10\lambda'_{ik})^{m_{ik}}e^{-10\lambda'_{ik}}}{\frac{1}{\theta_\lambda^{(\kappa_\lambda-1)}\Gamma(\kappa_\lambda)}\exp\left(\frac{-\lambda_{ik}}{\theta_\lambda}\right)\lambda_{ik}^{(\kappa_\lambda-1)}(10\lambda_{ik})^{m_{ik}}e^{-10\lambda_{ik}}}$$
$$\times\frac{\prod_{j=1}^{D}\frac{[\lambda'_{ik}(1-p_{ijk})]^{n_{ijk}}\,e^{-\lambda'_{ik}(1-p_{ijk})}}{n_{ijk}!}}{\prod_{j=1}^{D}\frac{[\lambda_{ik}(1-p_{ijk})]^{n_{ijk}}\,e^{-\lambda_{ik}(1-p_{ijk})}}{n_{ijk}!}}$$

Taking logs

$$\begin{aligned}
\log\frac{\pi(x')}{\pi(x)} &= (\lambda_{ik}-\lambda'_{ik})\frac{1}{\theta_\lambda}+(\kappa_\lambda-1)(\log\lambda'_{ik}-\log\lambda_{ik}) \\
&\quad +m_{ik}[\log(10\lambda'_{ik})-\log(10\lambda_{ik})]+10(\lambda_{ik}-\lambda'_{ik}) \\
&\quad +\sum_{j=1}^{D} n_{ijk}\log\left[\frac{\lambda'_{ik}(1-p_{ijk})}{\lambda'_{ik}(1-p_{ijk})}\right]-(\lambda'_{ik}-\lambda_{ik})(1-p_{ijk})A \\
&= \left[\frac{1}{\theta_\lambda}+10+\sum_{j=1}^{D}(1-p_{ijk})\right](\lambda_{ik}-\lambda'_{ik}) \\
&\quad +\left[\kappa_\lambda-1+m_{ijk}+\sum_{j=1}^{D} n_{ijk}\right]\log\frac{\lambda'_{ik}}{\lambda_{ik}}
\end{aligned}$$

The transition probabilities are:

$$\frac{q(\lambda'_{ik},\lambda_{ik})}{q(\lambda_{ik},\lambda'_{ik})}=\frac{\Phi\left(\frac{\lambda_{ik}}{s}\right)}{\Phi\left(\frac{\lambda_{ik}}{s}\right)}$$

Where s is the standard deviation of the Normal kernel used to generate proposals. The acceptance probability α is calculated as previously described.

5.6.13 Acceptance probability for updating n_{ijk}

The prior conditional distribution for N_{ijk} is given as $N_{ijk} \sim \text{Pois}((1-p_{ijk}\lambda_{ik})$. The ratio of likelihoods is calculated as follows:

$$\begin{aligned}
\frac{\pi(x')}{\pi(x)} &= \frac{f_{N|\lambda,B,X,X^*}(n'_{ijk}|\lambda_{ik},b_i,x_i^*,x_{ij})f_{Y|N,N0}(y_{ijk}|n'_{ijk},n0_{ik})}{f_{N|\lambda,B,X,X^*}(n_{ijk}|\lambda_{ik},b_i,x_i^*,x_{ij})f_{Y|N,N0}(y_{ijk}|n_{ijk},n0_{ik})} \\
&= \frac{\frac{[\lambda_{ik}(1-p_{ijk})]^{n'_{ijk}}}{n'_{ijk}!}\exp\left(-\lambda_{ik}(1-p_{ijk})\right)f_{Y|N,N0}(y_{ijk}|n'_{ijk},n0_{ik})}{\frac{[\lambda_{ik}(1-p_{ijk})]^{n_{ijk}}}{n_{ijk}!}\exp\left(-\lambda_{ik}(1-p_{ijk})\right)f_{Y|N,N0}(y_{ijk}|n_{ijk},n0_{ik})}
\end{aligned}$$

if $n'_{ijk} > n_{ijk}$:

$$\frac{\pi(x')}{\pi(x)} = \frac{[\lambda_{ik}(1-p_{ijk})]^{(n'_{ijk}-n_{ijk})}}{\prod_{q=n_{ijk}+1}^{n'_{ijk}} q} \frac{f_{Y|N,N0}(y_{ijk}|n'_{ijk}, n0_{ik})}{f_{Y|N,N0}(y_{ijk}|n_{ijk}, n0_{ik})}$$

if $n_{ijk} > n'_{ijk}$:

$$\frac{\pi(x')}{\pi(x)} = [\lambda_{ik}(1-p_{ijk})]^{(n'_{ijk}-n_{ijk})} \prod_{q=n'_{ijk}+1}^{n_{ijk}} q \times \frac{f_{Y|N,N0}(y_{ijk}|n'_{ijk}, n0_{ik})}{f_{Y|N,N0}(y_{ijk}|n_{ijk}, n0_{ik})}$$

The value for $f_{Y|N,N0}(y_{ijk}|n'_{ijk}, n0_{ik}$ may be zero if n'_{ijk} falls outside the valid range given y_{ijk} and $n0_k$. In this case the proposed value, n'_{ijk} is not adopted.

Transition probabilities are calculated as follows: If $n_{ijk} \pm k$, where k refers to the maximum jump size permitted, falls within the legitimate range then the proposal probabilities are uniformly distributed over $(n_{ijk}-k), ..., (n_{ijk}-1), (n_{ijk}+1), ..., (n_{ijk}+k)$. Similarly for n'_{ijk}. If this is the case for both n_{ijk} and n'_{ijk} then $q(n'_{ijk}, n_{ijk}) = q(n_{ijk}, n'_{ijk}) = 1/2k$, and so the ratio of transition probabilities is 1.

However, if the transition probabilities cease to be $1/2k$ owing to range constraints (as described in page 49) then this ratio is no longer 1 and must be calculated.

5.6.14 Acceptance probability for updating m_{ik}

The conditional probability for m_{ik} is Pois$(10\lambda_{ik})$. The ratio of likelihoods is calculated as follows:

$$\frac{\pi(x')}{\pi(x)} = \frac{f_{M|\lambda}(m'_{ik}|\lambda_{ik}) \prod_{j=1}^{D} f_{Y|N,N0}(y_{ijk}|n_{ijk}, m'_{ik})}{f_{M|\lambda}(m_{ik}|\lambda_{ik}) \prod_{j=1}^{D} f_{Y|N,N0}(y_{ijk}|n_{ijk}, m_{ik})}$$

$$= \frac{\frac{(10\lambda_{ik})^{m'_{ik}}}{m'_{ik}!}\exp(-10\lambda_{ik})f_{Y|N,M}(y_{ijk}|n_{ijk},m'_{ik})}{\frac{(10\lambda_{ik})^{m_{ik}}}{m_{ik}!}\exp(-10\lambda_{ik})f_{Y|N,M}(y_{ijk}|n_{ijk},m_{ik})}$$

if $m'_{ik} > m_{ik}$:

$$\frac{\pi(x')}{\pi(x)} = \frac{(10\lambda_{ik})^{m'_{ik}-m_{ik}}}{\prod_{q=0}^{m'_{ik}-(m_{ik}+1)}(m'_{ik}-q)} \times \prod_{j=1}^{D}\frac{f_{Y|N,M}(y_{ijk}|n_{ijk},m'_{ik})}{f_{Y|N,m}(y_{ijk}|n_{ijk},m_{ik})}$$

if $m_{ik} > m'_{ik}$:

$$\frac{\pi(x')}{\pi(x)} = (10\lambda_{ik})^{m'_{ik}-m_{ik}} \times \prod_{q=0}^{m_{ik}-(m'_{ik}+1)}(m'_{ik}-q) \times \prod_{j=1}^{D}\frac{f_{Y|N,M}(y_{ijk}|n_{ijk},m'_{ik})}{f_{Y|N,M}(y_{ijk}|n_{ijk},m_{ik})}$$

The transition probabilities are calculated in a similar manner to that in the updating of n_{ijk}.

5.6.15 Acceptance probability for updating y_{ijk}

The category of surviving cells has been given a categorical distribution dependent on $n0_{ik}$ and n_{ijk}. Details of calculating $\mathrm{P}(Y_{ijk} = y_{ijk}|N_{ijk} = n_{ijk}, N0_{ik} = n0_{ik})$ are given in Table 4.4.

$$\frac{\pi(y'_{ijk})}{\pi(y_{ijk})} = \frac{f_{Y_{ijk}|N_{ijk},N0_{ik}}(y'_{ijk}|n_{ijk},n0_{ik})}{f_{Y_{ijk}|N_{ijk},N0_{ik}}(y_{ijk}|n_{ijk},n0_{ik})}$$

Transition probabilities are calculated in a similar manner to that of m_{ik} and n_{ijk}.

5.7 Running the Markov Chain

Having obtained all the formulae required in order to perform updates in the Markov chain, these need to be implemented efficiently to try and reduce computing time.

Also, given the volume of data to be analysed, we wish to reach convergence as soon as possible. This involves monitoring acceptance rates and choosing suitable sized jumps to try and ensure the chain moves freely without getting stuck and that it has reached all parts of its posterior density. Before any reliable inferences can be drawn we have to be sure the chain has converged. This is done using convergence diagnostics within the package CODA.

This is dealt with in the following Chapter.

Chapter 6

MCMC Convergence and Validation

There are several important implementation issues associated with MCMC methods. The first involves the choice of sampler: we have chosen the Metropolis-Hastings algorithm because of the type of distributions being used in the model and the convenience of implementation this method offers when dealing with non-standard and non-conjugate distributions. Other important points to deal with include the choice of starting values, the length of the burn-in, acceptance rates of proposal values, and the number of iterations required in order to produce reliable estimates.

These factors are inter-related and all have some effect on the rate at which convergence occurs. Whilst having over dispersed starting values may delay convergence, they are a useful tool in determining convergence by comparing the results of running more than one chain (Gelman & Rubin 1992).

Various methods have been proposed to assess convergence and some of these which have been implemented in CODA (Best *et al.* 1995) were used to test some of the Markov chains produced. These techniques are looked at in more detail in Section 6.3. Brooks & Roberts (1999) and Cowles & Carlin (1996) give a comprehensive overview of other diagnostic methods in use and a comparative review of their applicability and ease of use.

Given the volume of data to be processed and hence the number of calculations involved, efforts are made to ensure the methods employed are efficient and that the chain mixes well to accelerate convergence.

6.1 Starting values

Whilst starting values are theoretically unimportant as, given any value, the chain will converge eventually, they may affect time to convergence (for example, see Raftery & Lewis 1992a). Thus it is desirable to have values which are not so far away from the posterior distribution that this adversely affects the rate at which the chain achieves convergence. Since the prior distribution should support the posterior probability distribution, using starting values generated randomly from the prior distribution, although diffuse, should not affect convergence.

This issue is usually a minor one with the continuous variables in this particular problem, but care needs to be taken with values which have range constraints. For example, when initialising the Y_{ijk}s, if the control value, $N0_{ik}$, is small and an inappropriate starting value is used, the chain may become stuck since the posterior distribution no longer assigns any probability to neighbouring categories.

This became evident with one patient with a particularly small control value of $n0_{30,1} = n0_{30,3} = 6$ and a set of cell survival categories ($Y_{30,j,k}$) with missing results as follows:

Dose		1	2	3	4	5
Slide 1,	*Counter A*	8	8	?	?	?
	Counter B	7	7	5	1	1
Slide 2,	*Counter A*	8	8	?	?	?
	Counter B	8	8	5	1	1

When the missing categories are updated, it is impossible to make a jump to category 1 to any of the others or *vice versa* as the data do not support any probability for the intermediate categories 2 and 3.

Category Y_{ijk}	Proportion surviving	Possible $N_{i,j,k}$ values
8	60%+	5, 6, 7, ...
7	45-100%	3,4,5,6
6	30-60%	2,3
5	15-45%	1,2
4	10-25%	1
3	2-12%	-
2	0.1-8%	-
1	0-2%	0

Table 6.1: *Possible values for surviving cells, N_{ijk}, for each category of proportion surviving*

In rare circumstances, such as this example, the chain *may* become reducible since the $n0_{ij}$ is small and n_{ijk} can take integer values only, hence not all categories of y_{ijk} are possible, see Table 6.1.

Under such circumstances, the updating scheme used to generate proposals prohibits movement between some categories as only steps of size ± 1 can be made when updating Y_{ijk}. This may be circumvented by changing the updating scheme to generate proposals over the whole range of categories rather than just the neighbouring ones. However, this would be rather wasteful as the proposals would be rejected most of the time.

The main problems with reducibility here is the inability for the chain to move from invalid or improbable categories to areas with of high probability if the intervening categories have no associated probability. Hence, in this particular example it is important that starting values for the missing Y_{ijk} categories are valid given $n0_{ik}$ and n_{ijk} so that any possible problems caused by reducibility are minimised. The chain may yield unreliable estimates otherwise as any proposed value for Y_{ijk} will be rejected since they will all be invalid. This prompted the initialisation scheme as described in Section 5.5 rather than simply assigning the category at the previous dose level for missing values.

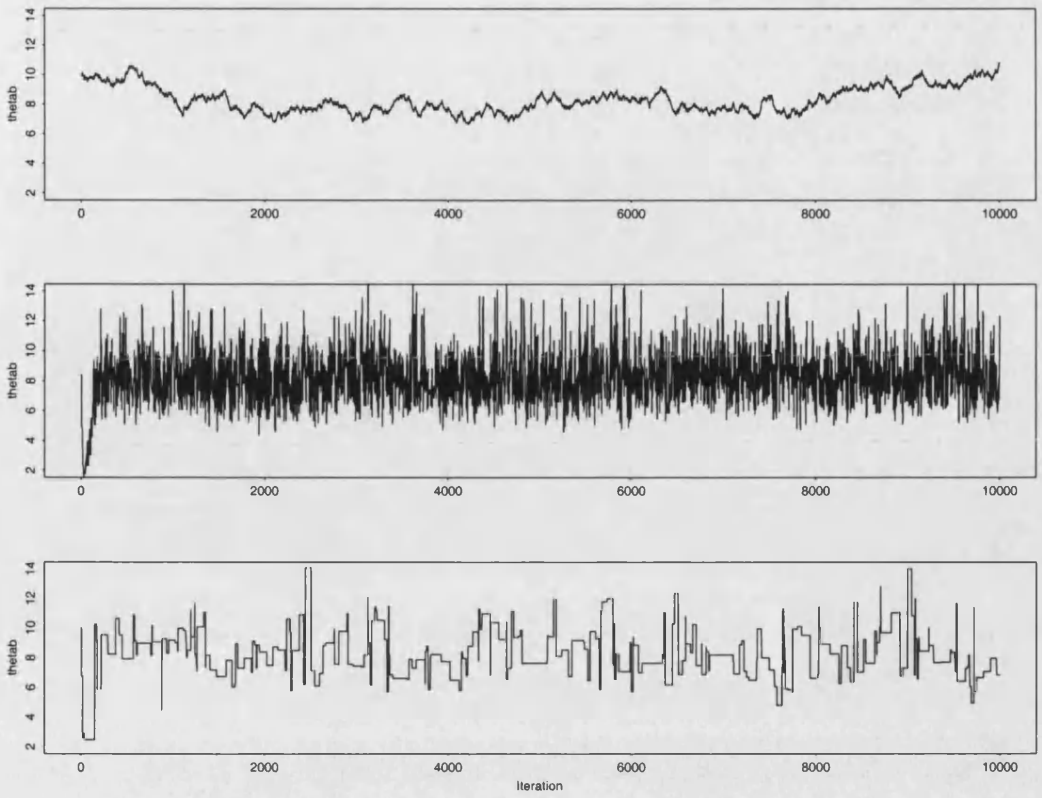

Figure 6-1: *Trace of Drug=VO, variable=θ_b.*

6.2 Acceptance rates

As stated in Section 5.6.1 the rate at which proposals are accepted also affects the time taken to convergence. If the proposals are too small, the chain moves very slowly through the target distribution. If proposals are too extreme then they are rarely accepted.

Figure 6-1 shows three example plots constructed using different standard deviations for the proposal kernel used for updating variable θ_b. The acceptance rates differ greatly depending on the standard deviation used: from top to bottom the acceptance rates for the traces are: a) 99%, b) 50% and c) 2.5%. If the standard deviation of the Normal kernel used to generate proposals is too small then the proposals are nearly always accepted and the chain "snakes" up and down and convergence is slow, as in the first trace shown. If the jumps are too

big then the chain may get stuck for periods of time and the trace looks "blocky" (last trace). See Section 5.6.1 for a discussion of desirable acceptance rates.

6.3 Convergence diagnostics

Convergence diagnostics are designed to assess how long to run a Markov Chain in order to obtain reliable observations from the stationary distribution.

Gelfand *et al.* (1990) broached the subject with what is now known as their *Thick pen technique.* This requires running a Markov chain m times, each of length t. Density estimates of the resulting distributions are drawn. For fixed m, the authors increase t and overlay plots of the resulting density estimates. If they coincide to within a *thick pen's width* then they can be said to have converged by time t. For example, Figure 6-2 shows the resulting density estimates from running a Markov chain from different starting values. Despite including any burn-in period, this plot shows good agreement of the estimates.

However, this is a time consuming and subjective exercise if the problem is other than trivial and the authors therefore advocate the development of easily implemented automated diagnostics.

Opinion is divided over whether it is best to use one long chain or several short ones, both cases have merits and disadvantages. Some authors advocate running multiple chains, for example Gelman and Rubin (1992), whereas others prefer the use of single chains e.g. Heidelberger and Welch (1983), Raftery and Lewis (1992b). Whilst multiple chains have the benefit of using different starting points from which to assess convergence later on, single chains are said to be less vulnerable to initialisation bias (Heidelberger and Welch, 1983). Single run methods are also more efficient since a burn-in period is required only once and is hence less wasteful of computer time. The problem of how long to run the simulation is further complicated with multiple chains by having more than one chain for which to assess convergence.

There are now many automated convergence diagnostics which have been developed but none are really foolproof (Brooks & Roberts 1999a). Care must be

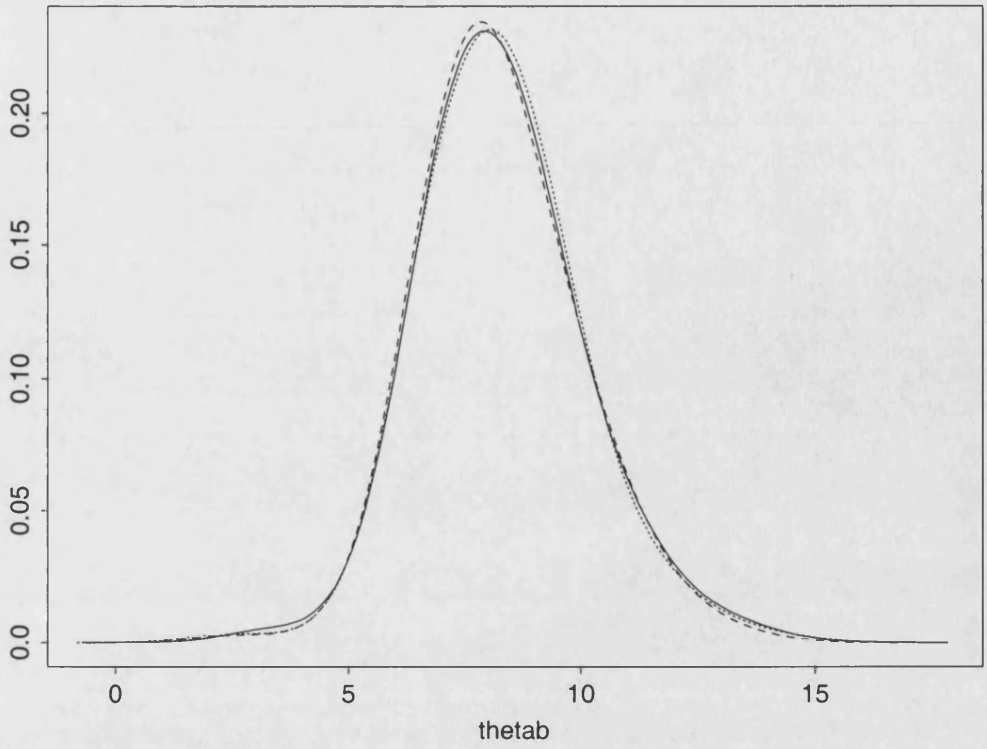

Figure 6-2: *Density estimates of the posterior distribution for θ_b from chains with different starting points, drug=VO*

taken as convergence may often be wrongly diagnosed (Roberts 1992). One of the problems of stationarity tests is the lack of power to detect an initial transient phase when the run length is shorter than that of the transient. However, despite their current shortcomings they are useful and a combination of both plots and diagnostics usually gives a good indication of convergence.

6.4 CODA

Some of the more easily implementable and therefore popular convergence diagnostics are readily available in the package CODA. (Best *et al.* 1995). CODA consists of a set of functions which run within the S-Plus environment. It was originally conceived as an output processor for the BUGS software (Spiegelhalter *et al.* 1994) to assess convergence of the chains produced. On a practical note, since BUGS is not used to analyse this problem, output must conform to the specific format produced by BUGS in order to be able to use CODA. The authors of CODA have implemented a range of convergence diagnostics in a convenient form, detailed below. We have used CODA to test a variety of patient/drug combinations: to test all variables would be prohibitive given the more than 2,000,000 analyses on the individual chains from each drug/patient/variable combination which would be required.

The examples given below illustrate the convergence diagnostics as implemented in CODA and refer the resultant chains from the drug= V0 and patient = 70.

6.4.1 Graphical output

Figure 6-3 shows plots of the traces for the variables examined for drug=VO recorded for a particular patient together with the corresponding density estimates. Where variables are constrained to be positive, *reflection* (Silverman, 1986) has been used when constructing these estimates. The traces appear to indicate that the chains have all reached their stationary distributions well within the burn-in period initially allowed despite the quite distinct starting values generated. All the traces show a rapid fluctuation about their stationary distribution indicating that the chains have been mixing well. The initial values and those generated during the burn-in period appear in the tails of the density estimates of the posterior distributions. These transient values have negligible effect when constructing the density estimates.

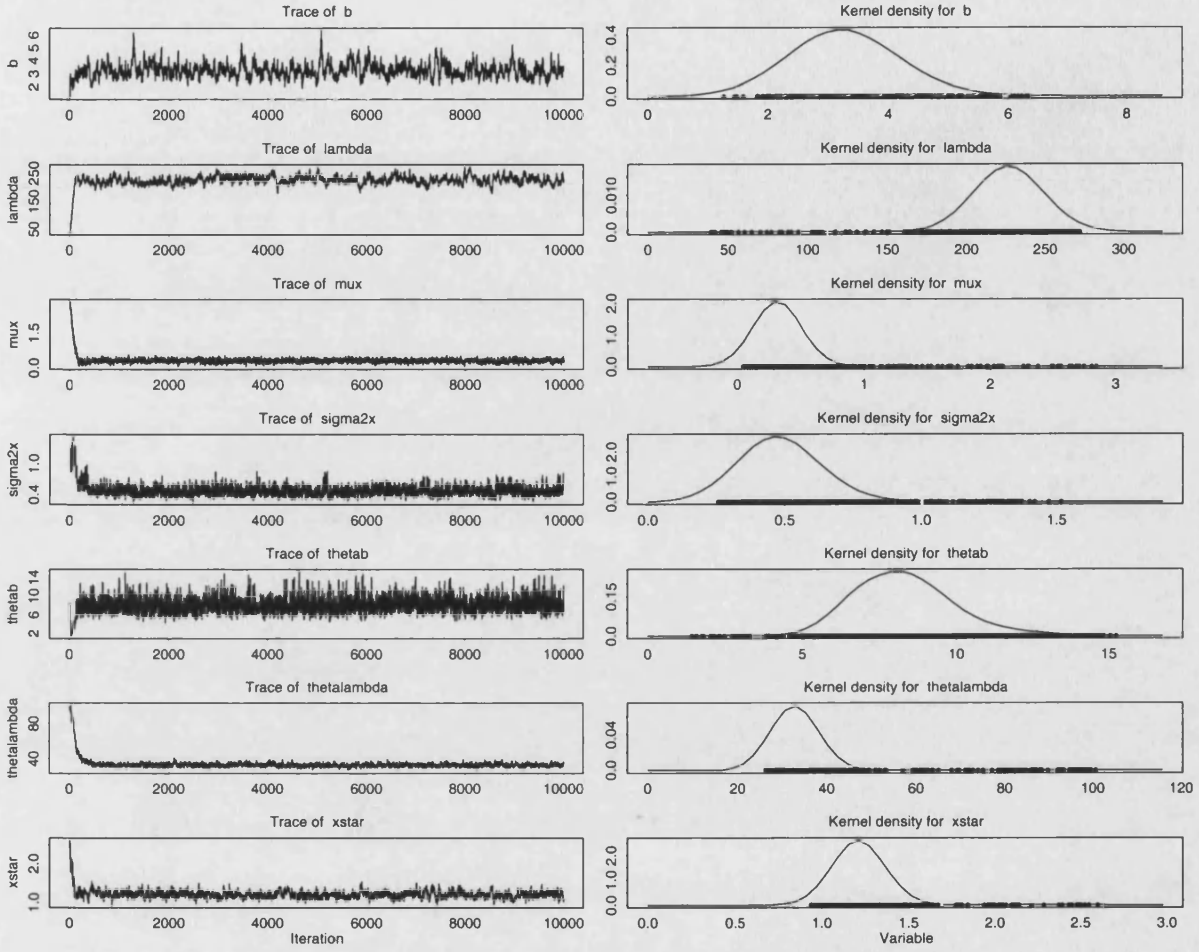

Figure 6-3: *Traces of variables for Drug=VO, patient=70.*

6.4.2 Geweke's convergence diagnostic

Geweke (1992) proposes a diagnostic using time series methods. Given the sequence $\{G(j), j = 1, 2, ...\}$, say, a comparison of values early in the sequence with those late in the sequence is likely to reveal failure of convergence. In order to perform this comparison we need an estimate of the variance.

The method employs standard techniques from Spectral Analysis in order to gain variance estimates via the spectral density, $S(\omega)$.

The spectral representation of a stationary process, $\{G(t), t = 0, \pm 1, ...\}$, essentially decomposes $\{G(t)\}$ into a sum of sinusoidal components with uncorrelated random coefficients (Brockwell & Davies, 1991). In conjunction with this decomposition there is a corresponding decomposition into sinusoids of the autocovariance function $\gamma(k)$ of $\{G(t)\}$.

$$\begin{aligned} \gamma(k) &= E\{[G(t) - \mu(t)][G(t+k) - \mu(t+k)]\} \\ &= \int_0^\pi \cos(\omega k) \mathrm{d}F(\omega) \quad \text{for } k = 0, 1, 2, ... \end{aligned}$$

Where the derivative, $\mathrm{d}F(\omega)$, exists, we obtain the Spectral Density function $S(\omega)$ so that

$$S(\omega) = \frac{\mathrm{d}F(\omega)}{\mathrm{d}\omega}.$$

The autocovariance function can thus be re-expressed as:

$$\gamma(k) = \int_0^\pi \cos(\omega k) S(\omega) \mathrm{d}\omega.$$

Putting $k = 0$ we then have:

$$\gamma(0) = \sigma_G^2 = \int_0^\pi S(\omega) \mathrm{d}\omega = F(\pi).$$

The variance function thus is a special case of the autocovariance function when

the lag $k = 0$ (Chatfield, 1989). The autocovariance function and the spectral density function are equivalent ways of describing a stationary stochastic process, they contain the same information expressed in different ways.

The spectral density function is given as:

$$S(\omega) = 2\left[1 + 2\sum_{k=1}^{\infty} \rho_k \cos(2\pi\omega k)\right]$$

where ρ_k is the autocorrelation at lag k and ω is the frequency.

In practice a smoothed version is used since the estimated sample spectrum fluctuates wildly about the theoretical spectrum and is inconsistent as the series length increases (Green & Han 1999). This is given by:

$$\hat{S}(\omega) = 2\left[\hat{\gamma}(0) + 2\sum_{k=1}^{N-1} \lambda_k \hat{\gamma}(k) \cos(2\pi\omega k)\right] \tag{6.1}$$

where $\hat{\gamma}(k)$ is the estimated autocovariance at lag k and the λ_k are suitably chosen weights called a *lag* window. The λ_k decrease to 0 as $k \to \pm\infty$.

The sample spectrum, Equation 6.1, is the Fourier cosine transformation of the estimate of the autocovariance function (Box *et al.*, 1994)

The integrated autocorrelation time is simply 2π times the *normalised* spectral density function evaluated at frequency $\omega = 0$, (Priestly, 1981, Chapter 4). The normalised spectral density function is simply given as

$$f^*(\omega) = \frac{S(\omega)}{\sigma_G^2}.$$

The asymptotic variance for $\overline{g} = \frac{1}{n}\sum_{i=1}^{n} G(j)$ is given as the spectral density at frequency 0: $\hat{S}_G(0)/n$.

Having obtained a formula to calculate an estimate of the variance, the convergence diagnostic is calculated by dividing the chain into two portions consisting of the first 10% and last 50% after discarding a burn-in phase and then performing a comparison.

We can now calculate Geweke's convergence diagnostic as follows. Let

$$\overline{g}_p^A = \frac{1}{p_A}\sum_{j=1}^{p_A} G(j) \quad \text{and} \quad \overline{g}_p^B = \frac{1}{p_B}\sum_{j=p^*}^{p_B} G(j)$$

where $(p^* = p - p_B + 1)$. Also let $\hat{S}_G^A(0)$ and $\hat{S}_G^B(0)$ denote the consistent spectral estimates for $\{G(j), j = 1, ..., p_A\}$ and $\{G(j), j = 1, ..., p_B\}$ respectively.

The asymptotic variance of $\overline{g}_p^A$ is $p^{-1}S_G^A(0)$, similarly for $\overline{g}_p^B$. Geweke (1992) reports calculating $\hat{S}_G(0)$ using a Daniell window of width $2\pi/M, where M = (0.3p^{1/2})$.

Thus, we can derive an expression for the asymptotic standard error for the difference of the means and hence obtain the test statistic Z:

$$Z = \frac{(\overline{g}_p^A - \overline{g}_p^B)}{\left[\frac{1}{p_A}\hat{S}_G^A(0) + \frac{1}{p_B}\hat{S}_G^B(0)\right]^{1/2}}$$

As the length of the chain $n \to \infty$, the sampling distribution of $Z \to N(0,1)$ if the chain has converged. Large values of Z indicate that the chain has not fully converged early on.

CODA output for the Geweke convergence diagnostic

Table 6.2 shows the output produced by CODA. There is evidence to show that all chains apart from that for b have converged within the 10,000 iterations used since the Z-scores are reasonably small. However, the Z-score for b is fairly borderline using traditional cut-off values and running this diagnostic on the same patient/drug combination with different starting values yielded slightly different results and this time b yields a Z-score of -1.28 suggesting there is not a serious problem with convergence here.

This diagnostic appears to yield somewhat erratic results when tested on different portions of the same chain. Running the same chain for 50,000 iterations and

```
GEWEKE CONVERGENCE DIAGNOSTIC (Z-SCORE):
========================================

Iterations used = 1001:10000
Thinning interval = 1
Sample size per chain = 9000

Fraction in 1st window = 0.1
Fraction in 2nd window = 0.5

 +---------------+------------+
 | Variable      |   vo       |
 | ========      |   ==       |
 | b             |    1.990   |
 | lambda        |    0.473   |
 | mux           |    0.730   |
 | sigma2x       |    1.240   |
 | thetab        |   -0.213   |
 | thetalambda   |   -1.330   |
 | xstar         |   -1.590   |
 +---------------+------------+
```

Table 6.2: *CODA output for Geweke convergence diagnostic.*

testing for convergence after every 10,000 iterations gives the Z-scores in Table 6.3. Convergence is diagnosed differently dependent on the length of the segment examined, the Z-score varying substantially. Each chain segment has had at least one variable deemed to have "failed" the convergence test.

One documented disadvantage of Geweke's diagnostic is its sensitivity to the specification of the spectral window (Cowles & Carlin, 1996). The spectral window is essentially a weight function expressing the contribution of the spectral density function to the expectation of $\hat{S}(\omega)$ (Chatfield, 1989). Also, intuitively, one would expect that using a longer chain would increase the chances of convergence begin diagnosed, however, with large samples, results of hypothesis tests are often significant even when the effects are quite small (Chatfield, 1988, Chapter 7). As with all hypothesis testing, multiple tests increases the risk of a Type I error and the results from this diagnostic test must be used cautiously.

6.4.3 Gelman and Rubin convergence diagnostic

The diagnostic proposed by Gelman and Rubin (1992) is a method based on running more than one parallel chain, each started from different over-dispersed starting values with respect to the stationary distribution. Multimodal target

	Number of Iterations				
Variable	10,000	20,000	30,000	40,000	50,000
b	-1.99	-1.88	0.52	-3.29	-2.42
λ	0.47	-0.20	-1.36	-0.33	-0.33
μ_{x^*}	0.73	-0.32	-2.04	-1.71	-1.28
$\sigma^2_{x^*}$	-1.24	1.25	1.57	2.25	1.89
θ_b	-0.21	-2.38	-1.28	-1.24	-1.16
θ_λ	-1.33	0.70	0.52	-1.31	-1.56
x^*	-1.59	1.58	-1.05	-0.66	1.11

Table 6.3: *Z-scores from the Geweke convergence diagnostic from different sized chain portions.*

distributions can give iterative simulation algorithms serious problems because the chains may take a long time to move from the region of one mode to another and may not be long enough to make these moves. Running several chains can reduce this risk of not visiting all modes of the distribution. This approach also provides useful information about how well the chains are mixing by comparing characteristics from each chain produced. However, this method runs the risk of spending too long in such extreme modes with respect to the target distribution and results may be biased if the chain is run for a sufficiently long time.

The technique is based on a comparison of the within-chain and between-chain variances for each variable, similar to an analysis of variance.

This diagnostic involves running $m \geq 2$ sequences each of length $2n$ and having different starting points. The method disregards the first n iterations and uses only the last n. For any scalar function of interest $\theta(x)$, convergence is monitored by estimating the factor by which the scale of the posterior distribution for θ shrinks as n increases and the early over-dispersion is lost.

The estimated shrink factor, $\hat{R}$, is given as:

$$\sqrt{\hat{R}} = \sqrt{\left(\frac{n-1}{n} + \frac{m+1}{m\,n}\,\frac{B}{W}\right)\frac{df}{df-2}},$$

where B is the sample variance between means of the m chains, W is the average of the m within chain variances, and df is the degrees of freedom.

```
GELMAN AND RUBIN 50% AND 97.5% SHRINK FACTORS:
==============================================

Iterations used for diagnostic = 5001:10000
Thinning interval = 1
Sample size per chain = 10000

 +---------------+-----------------------------+
 | Variable      |   Point est. 97.5% quantile |
 | ========      |   =========  ============== |
 | b             |    1.01       1.03          |
 | lambda        |    1.01       1.05          |
 | mux           |    1.00       1.00          |
 | sigma2x       |    1.00       1.01          |
 | thetab        |    1.00       1.00          |
 | thetalambda   |    1.02       1.08          |
 | xstar         |    1.00       1.00          |
 +---------------+-----------------------------+
```

Table 6.4: *CODA output for Gelman and Rubin convergence diagnostic.*

The factor computed will eventually decline to 1 as $n \to \infty$ and the measure of convergence is the closeness of the estimated factor to 1. However, the authors give no indication of how close to 1 the shrink factor needs to be to be considered satisfactory.

CODA output for the Gelman & Rubin convergence diagnostic:

CODA reports the median and 97.5% quantiles of the sampling distribution for the scale parameter by which the marginal posterior distribution *may* be reduced if the chain were run to infinity. If the shrink factors for both these quantiles are close to 1.0, it may be inferred that effective convergence has occurred and the samples may be assumed to have arisen from the stationary distribution.

In Table 6.4 all the values are fairly close to 1 and we can conclude that there is no evidence that the chains have not converged. The plots in Figure 6-4 show that for all variables the median and 97.5% quantile for the shrink factors, except b, have stabilised to around 1.0 well before the first 1000 iterations. The slope variable b takes a little longer but the associated shrink factors are satisfactory. Despite the aforementioned problems associated with using multiple chains, this diagnostic yields consistent results when applied to different chains and thus appears to be fairly reliable when diagnosing convergence.

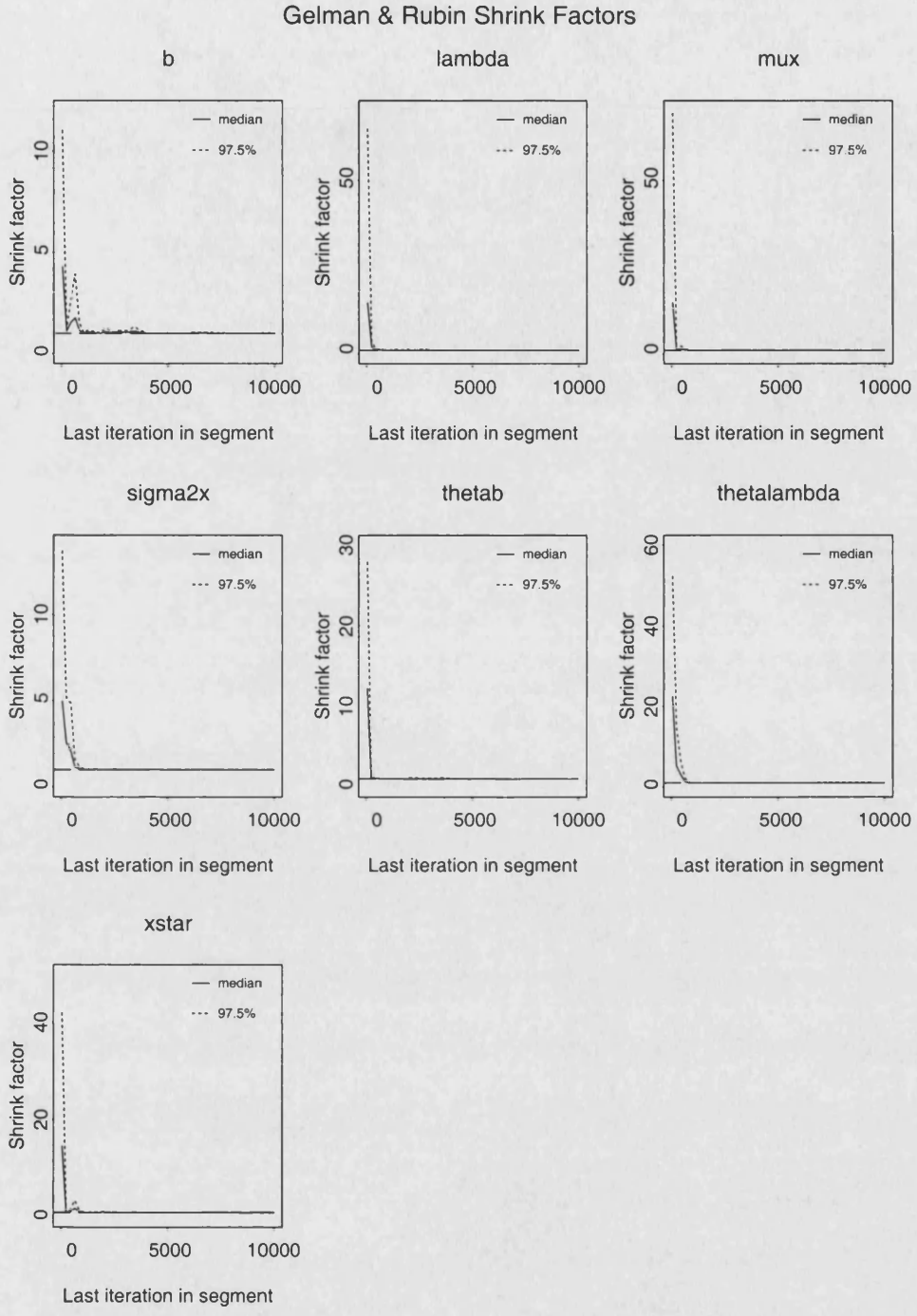

Figure 6-4: *Graphs showing Gelman & Rubin Shrink Factors.*

6.4.4 Raftery and Lewis convergence diagnostic

Raftery and Lewis (1992b) consider the problem of calculating the number of iterations necessary to estimate a posterior quantile from a single run of a Markov chain. For some function θ of a parameter (or set of parameters) X, we wish to estimate from the data the value, u, such that

$$P(\theta(X) \leq u) = q$$

for some stated value q so that the $P(\theta(X) \leq \hat{u})$ lies with $\pm r$ of the true value with probability p. The method involves first calculating the length of the burn-in period and secondly the number of further iterations, n, required to estimate the probability to within the required accuracy. They also suggest *thinning* the the chain by storing only every s^{th} iteration. The estimated number of iterations required for the desired accuracy actually increases when estimating quantiles nearer to the median compared to more extreme quantiles. If these quantiles are not of interest, this method must be used with caution (Brooks & Roberts, 1999a). In another paper, Brooks & Roberts (1999b) go on to suggest that rate of convergence is consistently overestimated and hence the suggested burn-in lengths are underestimated. They recommend using the method for a number of quantiles and to take the largest burn-in length, although if quantiles themselves are not of interest this method should not be relied upon to provide definitive answers.

CODA output for the Raftery and Lewis convergence diagnostic:

The output as given in Table 6.5 for precision ± 0.01 and probability of 97.5% shows the parameters required in order to obtain accurate estimates of the 2.5% quantile. If the samples in the chain were independent, then *Nmin* would be the minimum number of iterations required. Since the samples are not independent, the final column showing the dependence factor, we need to take a longer run length and then thin the chains by the factor shown in the first column. Thinning the chains reduces the autocorrelation but can been seen as wasteful. The higher the dependence factor, the more iterations are required. However, since we are

```
RAFTERY AND LEWIS CONVERGENCE DIAGNOSTIC:
=========================================

Iterations used = 1:10000
Thinning interval = 1
Sample size per chain = 20000

Quantile = 0.025
Accuracy = +/- 0.01
Probability = 0.95

Chain: vo
=========
```

Variable	Thin (k)	Burn-in (M)	Total (N)	Lower bound (Nmin)	Dependence factor (I)
b	7	49	12208	937	12.5
lambda	2	72	19362	937	14.5
mux	2	16	4230	937	5.2
sigma2x	1	13	3574	937	3.9
thetab	1	21	5846	937	0.5
thetalambda	1	18	4789	937	4.9
xstar	1	21	5632	937	6.9

Table 6.5: *CODA output for Raftery and Lewis convergence diagnostic.*

most interested in the LC90 and the output recommends taking all values for X^*, this does not affect the validity of our results. It can be seen that both b and λ need to run by longer than the run length given to achieve the specified accuracy with the desired probability. By relaxing these parameters to those given in the CODA manual (0.02 and 0.9 respectively) the run length now required ranges from a minimum of 640 for $\sigma^2_{x^*}$ to a maximum of 3470 for λ which is less than the 10,000 iterations that we used.

Critics of this method point out that the technique can produce variable estimates of the required number of iterations given different initial values for the same problem (Cowles & Carlin, 1996): applying this diagnostic to the same problem with different starting values yields wildly different results in the estimated recommended run length for b, see Table 6.6(a) and compare the values obtained with those in Table 6.5. The associated value for N is now given as over 400,000, although the other variables have recommended run lengths well within the number of iterations actually used. Reducing the initial accuracy to ± 0.02 reduces these values approximately fourfold (Table 6.6(b)). However, when we increase the run length and then re-apply the diagnostic the value for b decreases

Run length	(a)10,000		(b)10,000		(c)20,000	
Accuracy	±0.01		±0.02		±0.02	
	Burn-in	Total	Burn-in	Total	Burn-in	Total
Variable	M	N	M	N	M	N
b	1681	468,823	1681	116,467	159	10539
λ	38	9471	38	2397	56	3656
μ_{x^*}	13	3584	13	906	13	904
$\sigma^2_{x^*}$	11	3045	11	770	12	821
θ_b	33	9021	33	2280	18	1225
θ_λ	17	4516	17	1142	16	1079
x_*	21	5797	21	1465	30	2052

Table 6.6: *Raftery & Lewis diagnostic for differing accuracies and run-lengths*

dramatically to marginally over 10,000. The N for other variables also changes but their direction is not consistent, see Table 6.6(c).

Another criticism of using this method to determine run length is that sub-sampling from a chain produces a poorer estimate than when using the full chain. MacEachern & Berliner (1994) prove this and advise that thinning chains is inappropriate in most cases. Sub-sampling in this case is used to simplify the calculation of the diagnostic.

6.4.5 Heidelberger and Welch's convergence diagnostic

The method of Heidelberger and Welch (1983) is concerned with identifying transient phases in simulated sequences of discrete events using Brownian bridge theory. If we have a sequence $\{Y(j), j = 1, ..., n\}$ from a stationary process with spectral density $p(f)$, let

$$\begin{aligned} S_0 &= 0, \; S_n = \sum_{j=1}^{n} Y(j), \; n \geq 1 \\ \overline{Y} &= \frac{1}{n}\sum_{j=1}^{n} Y(j), \\ B_n(t) &= \left(S_{[nt]} - \frac{[nt]\overline{Y}}{np(0)} \right)^{1/2}, \; 0 \leq t \leq 1, \end{aligned}$$

where $[nt]$ indicates the greatest integer $\leq nt$. Then $B_n = \{B_n(t), 0 \leq t \leq 1\}$ converges in distribution to the Brownian bridge $B = \{B(t), 0 \leq t \leq 1\}$ as $n \to \infty$. So for large n, B_n is distributed approximately as the Brownian bridge B. The null hypothesis that the sampled values for a given variable form a stationary process is tested using the Cramer-von-Mises statistic:

$$\int_0^1 B_n(t)^2 dt = \text{CVM}(B_n).$$

This statistic has a known asymptotic distribution under stationarity and is sensitive to the existence of an initial transient. As $n \to \infty$ it converges in distribution to:

$$\int_0^1 B(t)^2 dt = \text{CVM}(B).$$

In practice, since $p(0)$, the spectral density at zero frequency is unknown, the approximation

$$\hat{B}_n(t) = \frac{S_{[nt]} - [nt]\overline{Y}}{(n\hat{p}(0))^{1/2}}, \quad 0 \leq t \leq 1,$$

is used instead.

If the null hypothesis is rejected, the first 10% of iterations are discarded. The process is repeated until either the the chain passes the stationarity test or 50% of the iterations have been discarded.

The authors of this diagnostic found that the stationarity test had little power to detect an initial transient when the run length was shorter than the extent of the initial transient.

CODA output for the Heidelberger and Welch convergence diagnostic:

The output produced by CODA in Table 6.7 reports the number of iterations to keep for each variable, i.e. those which are diagnosed to have arisen from a stationary process if the stationarity test has been passed. The halfwidth is calculated to be 1.96 $\times$ asymptotic standard error: if it is less than ϵ times the sample mean (where ϵ is a small fraction), then the retained sample is deemed to estimate the posterior mean with acceptable precision.

```
HEIDELBERGER AND WELCH STATIONARITY AND INTERVAL HALFWIDTH TESTS:
================================================================

Iterations used = 1001:10000
Thinning interval = 1
Sample size per chain = 9000

Chain: vo
=========
```

Variable	Stationarity test	# of iters. to keep	# of iters to discard	C-vonM stat.
b	passed	10000	0	0.314
lambda	passed	8000	2000	0.454
mux	passed	9000	1000	0.129
sigma2x	passed	9000	1000	0.194
thetab	passed	7000	3000	0.378
thetalambda	passed	9000	1000	0.385
xstar	passed	9000	1000	0.098

Variable	Halfwidth test	Mean	Halfwidth
b	passed	3.28	0.049
lambda	passed	228.0	1.330
mux	passed	0.32	0.004
sigma2x	passed	0.48	0.005
thetab	passed	8.42	0.869
thetalambda	passed	32.7	0.122
xstar	passed	1.21	0.007

Table 6.7: *CODA output for Heidelberger and Welch convergence diagnostic.*

All the variables in our example pass both the stationarity test and the halfwidth test. However, it suggests discarding more than the initial burn-in period allowed for both λ and θ_b.

6.4.6 Correlations

High cross correlations amongst parameters are associated with slow convergence and may indicate the need for re-parameterisation of the model (Best *et al.*, 1995). The cross-correlation matrix for the chosen example is shown in Table 6.8. All the values are fairly low showing no obvious problems.

Variable	b	λ	μ_{x^*}	$\sigma^2_{x^*}$	θ_b	θ_λ	x^*
b	1.000						
λ	-0.063	1.000					
μ_{x^*}	-0.017	0.024	1.000				
$\sigma^2_{x^*}$	-0.048	-0.005	0.005	1.000			
θ_b	0.035	-0.013	-0.029	-0.038	1.000		
θ_λ	0.002	0.024	-0.027	0.007	-0.025	1.000	
x^*	-0.419	-0.277	0.0516	0.022	0.001	-0.048	1.000

Table 6.8: *Table of cross correlations*

Autocorrelations

Autocorrelations measure the corrrelation between observations at different distances apart, k where $k = 1...N-1$. They can be useful in indicating the rate of convergence in a chain. Although short term correlation is to be expected even for a stationary chain this will get successively smaller for greater values of k. High long term correlations within chains indicate slow mixing and, usually, slow convergence. This is characterised by plots of traces which move slowly as opposed to showing more rapid fluctuations over the sample space. Very low or very high acceptance rates lead to highly correlated samples. One way of compensating for this is to *thin* the chain by taking every n^{th} observation and discarding the rest, however this has been shown to reduce the accuracy and so is rather wasteful. Müller (1993) suggests a scheme for monitoring the acceptance rates within the chain over the last 10 iterations and rescaling the size of jumps possible by $\pm 20\%$. This would not be appropriate for the discrete variables in this problem and the values proposed are very subjective with no account given of their suitability.

6.4.7 Summary of convergence diagnostics

As stated previously, no convergence diagnostic is foolproof and some can give varying results depending on which segment of the chain we look at, the run length used and the starting values used. Table 6.9 gives a summary of whether each technique diagnosed convergence or otherwise for the illustrated example.

Variable	Geweke	Gelman & Rubin	Raftery & Lewis (a)	(b)	Heidelberger & Welch
b	borderline	yes	borderline	yes	yes
λ	yes	yes	no	yes	yes
μ_{x^*}	yes	yes	yes	yes	yes
$\sigma^2_{x^*}$	yes	yes	yes	yes	yes
θ_b	yes	yes	yes	yes	yes
θ_λ	yes	yes	yes	yes	yes
x^*	yes	yes	yes	yes	yes

Table 6.9: *Comparison of convergence diagnostics*

The methods of both Geweke and Raftery & Lewis can give inconsistent results. Geweke's diagnostic appears to be over-sensitive and large run lengths appear to exacerbate the problem; this may be from underestimating the variance used to compute the Z-score. Raftery & Lewis's technique also varies depending on the chain length used, the same problem can yield very different answers using different starting points. The methods of Gelman & Rubin appear to yield consistent results although their interpretation is subjective and not straightforward.

Whilst these diagnostics are useful, they cannot be wholly relied on without further evidence of convergence or otherwise. A plot of the chain is often the most useful tool: although convergence is assessed visually, this is usually sufficient. The motivation behind automated convergence diagnostics appears to have been partly to remove the subjective element and speed. However, many techniques are slow even when automated and parameter choices still have to be made.

Having performed these diagnostic tests on a variety of drug/patient combinations, we are satisfied that the chains do converge within the length of burn in used, and often much before this. We are also satisfied with the run length of 10,000 iterations to yield reliable estimates. As we are especially interested in the posterior distribution of the LC90 and it passes all the diagnostic tests, its convergence is not affected by the other variables which take longer to converge.

Chapter 7

Results

Running the programs for each drug produces a wealth of information, of which the LC90 is the variable in which we are most interested and the most important from the view of the medical practitioner.

From the output of our analysis, we can learn much about each drug's behaviour from the output produced and how it can vary between different patients, drugs and tumour types. This information can be re-used to produce estimates of a new subject's LC90 from new assays without running the model using the whole dataset again, which can be very time consuming.

7.1 Chapter outline

We shall examine the results produced in a number of ways, a summary of which is given below:

- **Section 7.2 Population parameters**: This section deals with the main variables and how they differ between drugs.
- **Sections 7.3 and 7.4 Individual drugs**: The results from two drugs, ACD and IFN, are examined in detail here. The former is an example of a typically behaved drug and the latter is a drug for which it has been

difficult to obtain reliable estimates.

- **Section 7.5**: This section looks at the results for separate patients and how to interpret them.
- **Section 7.6 Future subjects**: The problem of how to analyse additional patients whose results were not included in the main analysis is looked at in this section.

7.2 Population parameters

A summary of the three main variables for each drug is given in Table 7.1. Whilst the underlying rate of cells per slide, λ_{ik}, does not differ greatly between drugs, the slope, b_i, and $\log_{10}$ LC90, x_i^* do as the drugs all behave in very different ways and the estimated dose response curves reflect this.

Since λ_{ik} does not depend on the drug behaviour, its variability between drugs is much reduced as, although this not a randomised study, the patients have not been allocated to assay drugs purely by tumour type or other systematic criteria related to the number of cells expected if left untreated. The average slope of the drug response curve varies from just over 2 to nearly 9, demonstrating differing changes in rates of survival versus $\log_{10}$ dose. The variability of slopes between patients within each drug is large, see, for example, Figure 7-1 which shows all dose response curves for one drug and illustrates this variation. The estimated $\log_{10}$ LC90s also show considerable variation both within and between drugs which can also be seen in Figure 7-1.

These results can be further broken down into the different tumour sample types. Figure 7-2 shows the estimated $\log_{10}$ LC90s according to the type of tumour. For viewing purposes, the vertical axis on the graph has been sorted according to the $\log_{10}$ LC90s for CLL tumour samples. The pattern clearly emerging shows the solid tumour clumps are the least responsive to nearly all drugs apart from PR, MAF and ASP. CLL is the most responsive sample type in approximately 75% of cases.

Drug	Number of Patients	Mean(λ)	sd(λ)	Mean(b)	sd(b)	Mean(x^*)	sd(x^*)
ACD	258	67.17	53.71	4.21	1.55	-0.99	0.91
ACL	245	71.23	50.37	7.43	1.97	-0.41	0.67
AMS	504	70.09	56.41	3.70	1.82	-0.51	1.36
AR	661	75.50	58.42	3.61	1.73	0.91	1.08
ASP	328	77.51	61.46	2.12	1.62	2.43	1.18
BC	297	63.62	53.70	7.38	2.17	1.61	0.66
BL	517	69.96	53.85	2.93	1.38	2.36	0.95
BU	155	71.96	58.26	3.65	1.58	2.66	0.89
CB	594	70.78	56.51	6.61	1.87	1.66	0.58
CC	465	71.03	57.79	7.15	2.10	1.44	0.67
CCA	169	79.89	59.81	2.91	1.37	1.80	0.98
CDA	876	83.29	57.50	3.77	1.58	-0.97	0.97
CHL	1048	80.35	59.39	5.49	1.86	0.74	0.80
CP	482	66.10	57.36	6.28	1.77	0.85	0.61
DC	720	83.87	58.07	1.55	1.41	1.80	1.31
DEX	369	72.83	54.16	2.03	1.57	2.00	1.19
DN	416	78.81	60.10	6.32	1.95	-0.65	0.71
DOX	1147	74.07	57.64	5.81	1.72	-0.44	0.79
EPI	1002	79.97	58.23	6.69	1.78	-0.38	0.71
FL	1052	81.53	59.73	3.88	1.51	-0.02	0.92
FU	331	53.93	48.75	3.61	1.62	2.27	0.89
GLA	111	71.96	66.93	5.50	2.42	3.39	0.91
HIF	553	70.68	57.25	8.90	2.17	0.81	0.74
HN	105	62.63	46.77	5.18	1.80	0.27	0.72
HU	189	65.92	54.50	4.83	2.01	3.25	0.74
IDA	432	73.27	59.40	5.42	1.79	-0.83	0.77
IFN	111	55.00	53.13	5.50	2.42	3.39	0.91
LP	514	70.94	59.29	5.09	1.81	0.82	0.76
MAF	1070	72.65	56.13	6.39	1.94	0.55	0.81
MEP	370	77.68	55.53	2.64	1.80	1.43	1.07
MMC	846	59.95	48.31	4.92	1.68	0.13	0.76
MP	406	76.07	59.09	4.29	1.87	2.88	0.84
MZN	545	71.50	57.62	4.98	1.99	-0.46	0.85
PR	975	76.31	57.18	2.51	1.74	1.94	1.06
TAX	63	65.78	48.28	5.55	1.87	2.11	0.80
TG	421	77.93	65.43	3.97	1.52	1.74	0.78
VB	417	67.55	54.68	4.06	1.81	0.70	0.95
VC	915	73.83	56.12	2.54	1.52	0.51	1.04
VD	254	67.43	59.34	3.21	1.65	1.05	1.01
VM	103	65.75	63.14	5.08	1.85	0.38	0.79
VO	84	65.01	50.82	3.21	1.54	0.31	0.81
VP	664	68.98	56.13	4.39	1.73	1.40	0.85

Table 7.1: *Posterior means and standard deviations of λ, b and x^* for all drugs over all sample types.*

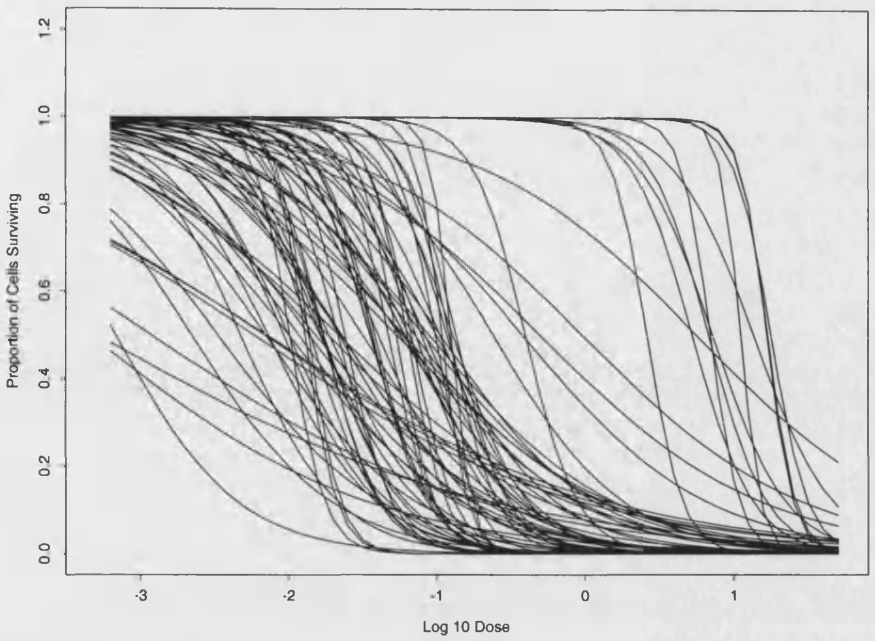

Figure 7-1: *Dose response curves for ACD showing all Solid samples.*

The overall trend in order of responsiveness is: CLL, Leukaemias & and Lymphonas, Normal, Solid tumours and Solid clumps. For many drugs, cell survival rates of solid tumours and solid clumps remained high at even the highest dose for a substantial proportion of subjects.

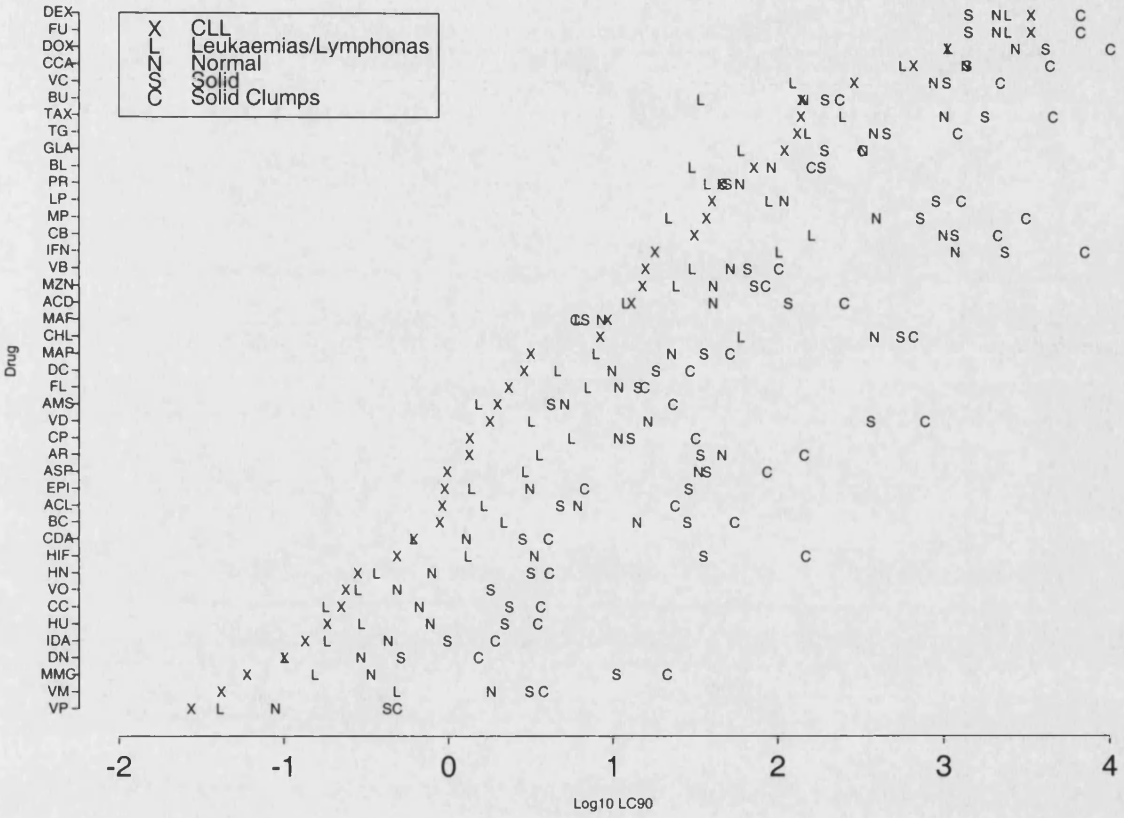

Figure 7-2: $\log_{10}$*LC90s for all drugs sorted by CLL results.*

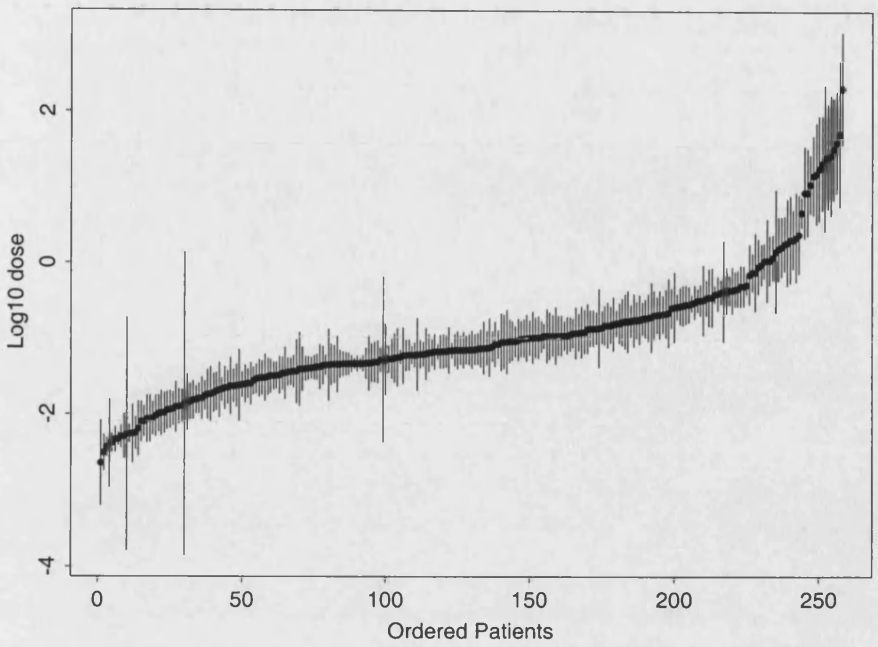

Figure 7-3: *Sorted* $\log_{10} LC90s$ *for ACD with 95% estimates.*

7.3 Individual drug: ACD

We shall examine the drug, ACD, initially. The estimated mean LC90s are shown in Figure 7-3 with corresponding 95% posterior intervals for each patient using the $\log_{10}$ Dose scale. From this graph, we can see that there are 3 patients with very wide intervals, patient numbers=67, 152, and 181. On further inspection, the latter two patients have very low cell densities on the control slide (1.3 and 0.1 respectively), and the remaining one has only one valid value for Y_{ijk}, the category of surviving cells. Both of these factors lead to imprecise estimates. With such small control values, a change in the estimated number of surviving cells, n_{ijk}, of just 1 cell, results in a large change in the proportion of surviving cells, which then gives wildly fluctuating LC90s. The lack of data for patient 67 yields a chain which has few constraints and moves about freely, again resulting in a very spread posterior distribution for the LC90.

Of more practical use to the clinician involved in selecting the drugs for treatment is the percentile of each patient's LC90 so assay results can be compared between

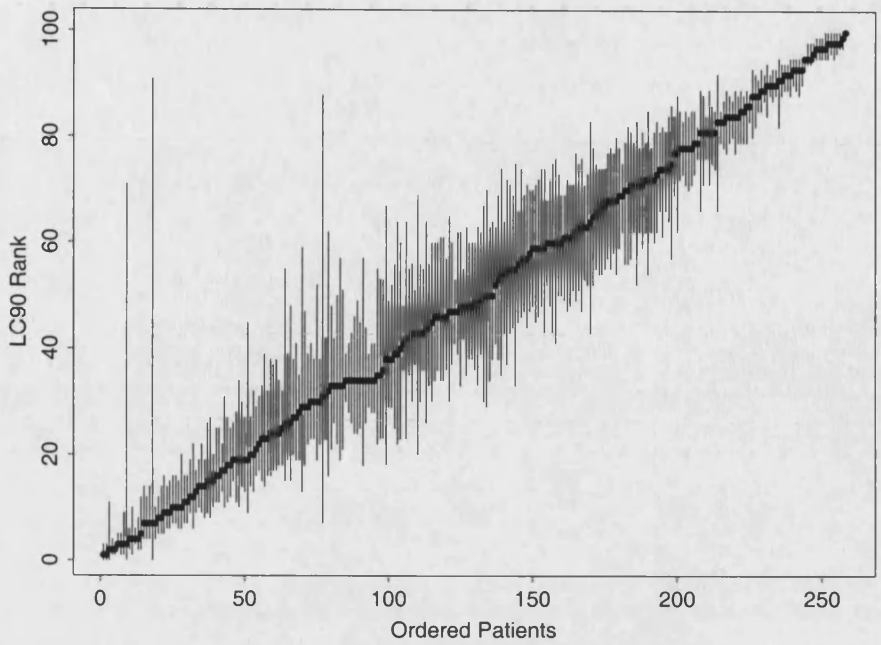

Figure 7-4: *Sorted LC90 Ranks for ACD with 5% and 95% percentiles.*

both drugs and patients.

The percentiles are calculated by ranking the current LC90 of each patient from $1, ..., p$ after each iteration of the MCMC algorithm, where p is the number of patients with assays for a particular drug. When the run has been completed the median rank achieved by each patient is noted, similarly for the 5% and 95% points which then give a 90% posterior interval for the rank of the patient's LC90 within all the other patients in the analysis. These results are then re-scaled from $1, ..., 100$ for ease of interpretation, see Figure 7-4.

Presenting results in this way is more natural for clinicians who can then report whether a patient is likely to respond well if their LC90 is in the top 10% say, of responses to that drug when compared to the other patients. This information can then be used in order to help prescribe a treatment using the indicated drugs or a combination of them. It must be noted, however, that a good relative response to a particular drug does not necessarily mean the patient will respond favourably. Some drugs may only be effective for a small proportion of cases,

and the dose required to achieve the LC90 for the next best response may be too toxic to tolerate despite having relatively high percentiles for the LC90. The percentiles must be used in conjunction with the estimated LC90 in order to ascertain the relative merits of particular drugs.

The use of percentiles however allows us to compare the relative merits of other drugs in order to get a comparison of potential response.

7.3.1 Dose response for ACD

We assumed the cell survival to have a logit relationship with dose since no cells were killed at zero dose and all cells were killed at a sufficiently large concentration. The model provides us with estimates for the slope of the dose response curve and also the $\log_{10}$ LC90 which gives us the location.

Figures 7-5 to 7-8 show the fitted dose response curves for single patients (patient numbers 15, 3, 69 and 179). These patients have been chosen to highlight some of the characteristics of the responses and also to show some unusual behaviour. On the graphs the vertical bars represent the recorded range of surviving cells; they have been staggered slightly for viewing purposes. There are up to 4 bars per dose depending how many sets of results were recorded and whether any of these values are missing. It is impossible to represent category 8 accurately as this is open ended and represents all cell survival rates over 60%. Where there are fewer cells on the control slides than the treated slides, the estimated proportion surviving may thus exceed 100%, we have therefore constrained the maximum range depicted to be 120% cell survival: this is only for pictorial convenience and the program has no such constraint.

To obtain the fitted cell survival curves we use the posterior mean of the $\log_{10}$ LC90, x_i^*, for the location of the curve and the posterior mean of b_i for the slope.

Figure 7-5 is a typical response curve with a comparatively steep slope ($b_{15} = 7.5$) and the LC90 being achieved between doses 3 and 4. It also shows the $Y_{15,3,k}$ values are discordant at dose level 3: the curve reflects this by passing between the categories. The other doses have good agreement for the cell survival categories

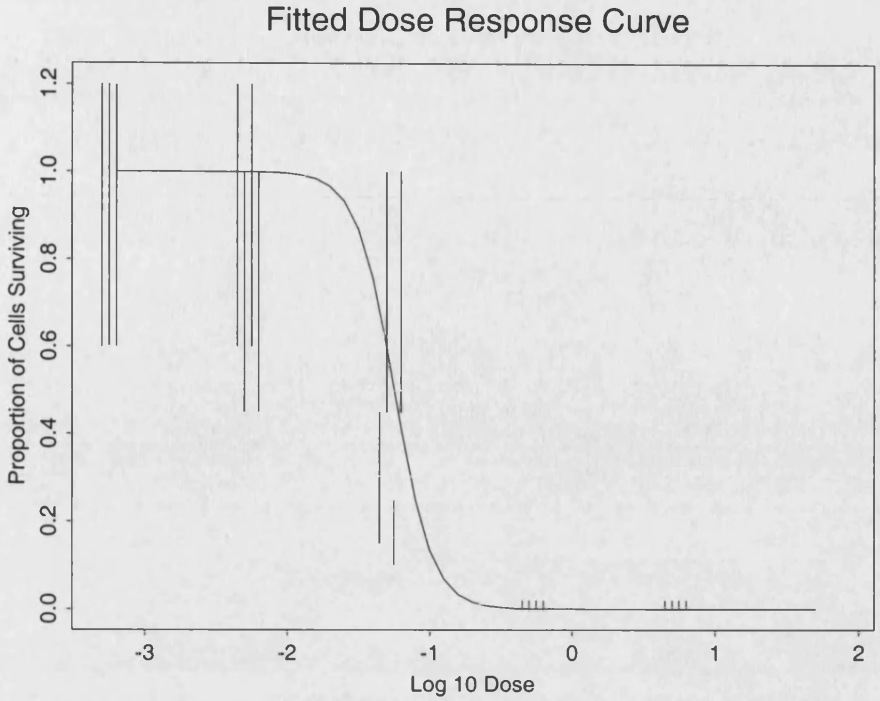

Figure 7-5: *ACD Dose response curve for Patient 15, drug=ACD*

with them all having the same recorded categories apart from dose 2 which differs by just 1 adjacent overlapping category, since one scientist was consistently recording lower values than the other.

The dose response curve for Patient 69 is shown in Figure 7-6. This patient has only two sets of results which differ widely for the two largest doses. The estimated $\log_{10}$ LC90 is 2.32, which is larger than the range of doses tested. The ranks achieved for this result range between 99–100% indicating a poor response is likely if this drug were used.

Figure 7-7 shows a patient (number=3) with a very sensitive *ex vivo* result. The estimated LC90 of 0.00229 μg/ml is achieved at a lower dose than any of the other patients. During the ranking process, the corresponding LC90 percentiles are 0-3% indicating that the result is reliable and the drug has a good chance of being effective.

Figure 7-8 shows the flattest dose response curve seen for this drug. The recorded categories do not agree well. One replicate actually shows an increase in the

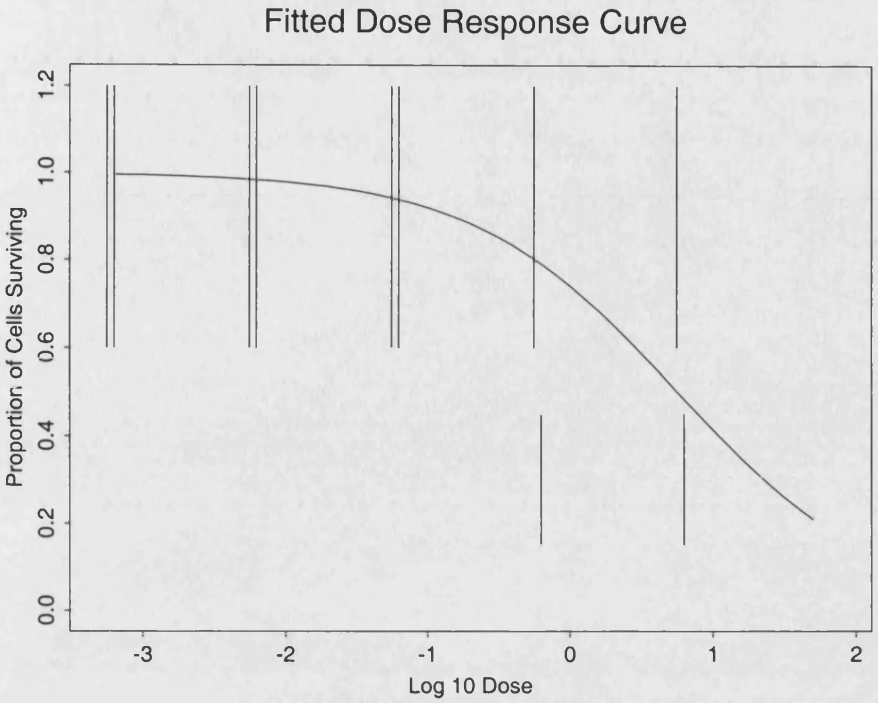

Figure 7-6: *ACD Dose response curve for Patient 69, drug=ACD*

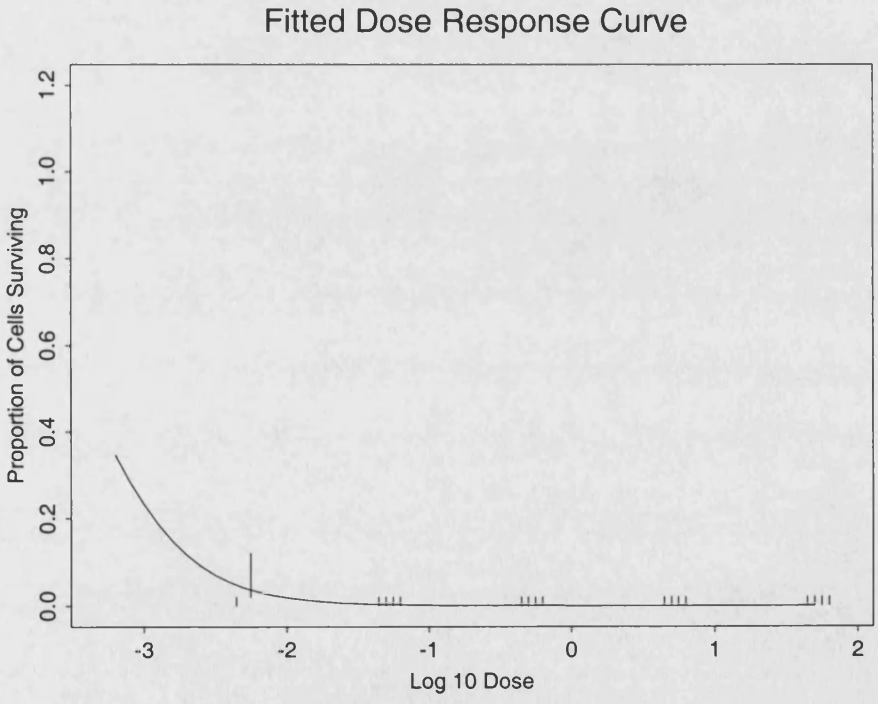

Figure 7-7: *ACD Dose response curve for Patient 3, drug=ACD*

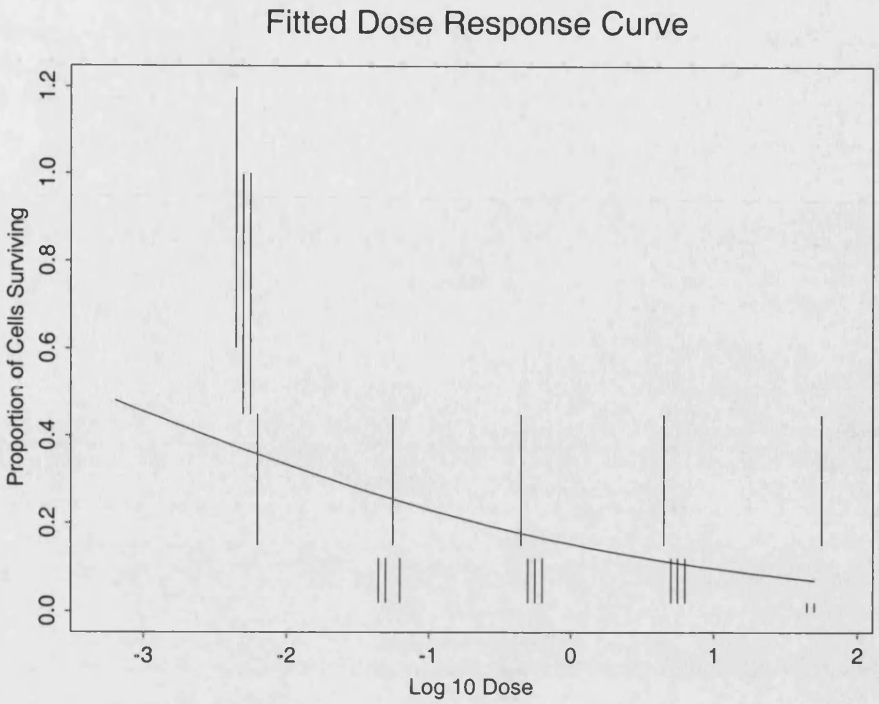

Figure 7-8: *ACD Dose response curve for Patient 179, drug=ACD*

proportion of cells surviving with dose. The fitted curve reflects this with a shallow gradient and an LC90 which is ranked between 94–98% in the patient population.

Although plotted on the same scale, these figures show a different range of doses used in the assays between the first two and latter two graphs. The latter graphs are in fact earlier assays and a maximum concentration of 50μg/ml was used. In assays performed later on, this was changed to 5μg/ml since the LC90 had been already been achieved at lower concentrations so the range of doses used was focused on a smaller area. The dilution factor remained the same regardless of the maximum concentration used.

Figure 7-9 shows all dose response curves for the drug ACD, broken down by tumour sample type. These graphs show steep slopes to be a common characteristic with slope gradients less then 1 occurring in only 6 of the 258 patients. The figure also reveals solid tumours to have far a more varied response curves and hence estimated LC90s than the other sample types. This variability

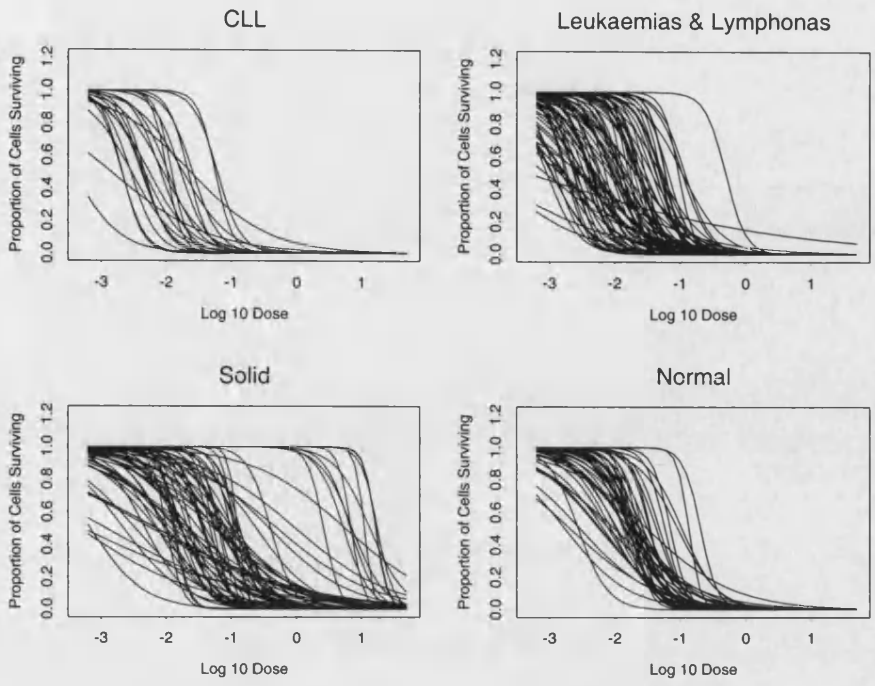

Figure 7-9: *ACD Dose response curves for all patients by tumour type*

may be due in part to many of these assays having smaller control values since cells are harder to distinguish and sometimes "clumps" are counted rather than individual cells, often leading to a very small $n0_{ik}$, reflected by low values for λ_i (Table 7.2).

The summary of the other important posterior distributions, broken down by tumour type, is shown in Table 7.2. Both the Solids and Solid Clump samples have lower underlying numbers of cells λ_i; they also show higher variability in slope, b_i and $\log_{10}$LC90, x_i^*. Their estimated LC90s are higher than other sample types showing a possible worse prognosis for Solid and SC tumours when treated with ACD.

In general, both slope and LC90, and hence the cell survival curves, vary considerably over all samples. This reflects the initial problem of choosing a drug for an individual patient since, even with the same disease type, patients respond very differently.

ACD is fairly representative of the drugs examined. The majority of patients

Tumour	x^*		b		λ	
Type	Mean	sd	Mean	sd	Mean	sd
Leukaemia & Lymphonas	-1.385	0.579	4.221	1.891	81.19	57.53
Normal	-1.053	0.348	4.093	2.062	88.36	61.20
Solid	-0.372	1.011	4.286	3.162	33.05	26.00
Solid Clumps	-0.311	0.871	3.982	3.092	40.15	29.83
CLL	-1.561	0.541	4.673	1.930	67.95	42.89

Table 7.2: *Posterior means and standard deviations for ACD.*

achieved the LC90 within the dose ranges tested. The least precise estimates are derived from patients with very little data or very small control values.

7.4 Drug IFN

In contrast to ACD, we shall also examine IFN as an example of a drug with data for which it has proved hard to obtain reliable estimates. Less than 10% of patients were assessed to have an LC90 within the range of doses tested: 0.4 – 102.4 μg/ml, (-0.39 – 2.02 on the $\log_{10}$ scale). Many assays show no recorded decrease in cell survival with increasing dose. Figures 7-10 and 7-11 show the plots of the LC90s with corresponding posterior intervals and the ranks achieved during the simulations.

Both plots reveal a few results with small confidence regions at a lower dose and the majority of patients with imprecise estimates at higher doses.

One patient has an uncharacteristically large posterior interval even though the LC90 is achieved at a relatively low dose (Patient 111). The reason for this is a very flat dose response curve, with slope $b_{111} = 0.8$ owing to missing data as well as recorded categories actually *increasing* with dose for one replicate, see Figure 7-12.

Figure 7-13 shows all dose response curves by disease type. They are very different from the drug ACL (Figure 7-9). There is a wide variety of both slope and location within all sample types. Many curves have very low gradients; these are associated with little or no decrease in surviving cells over the range tested:

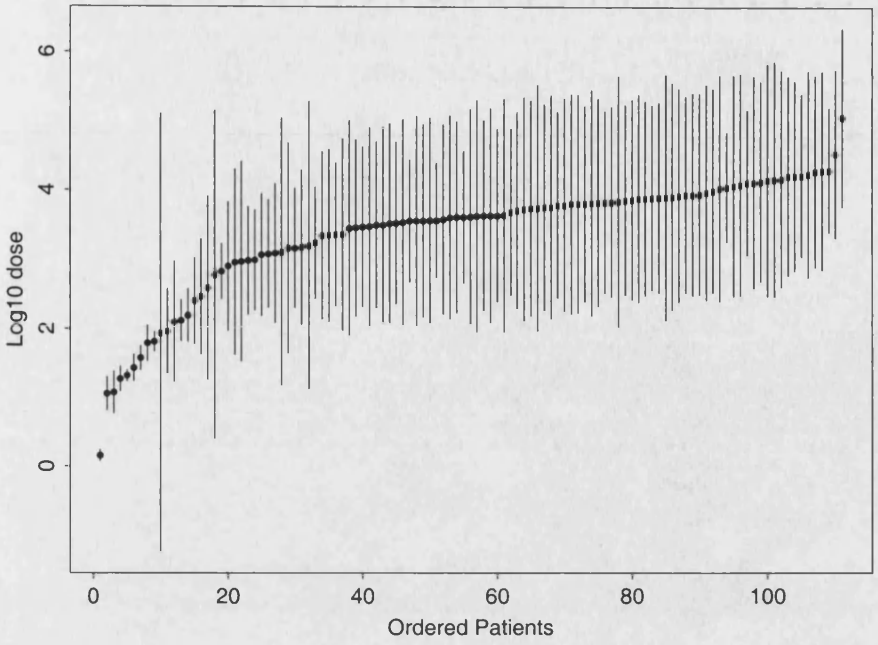

Figure 7-10: *Sorted* $\log_{10} LC90s$ *for IFN with 95% posterior intervals.*

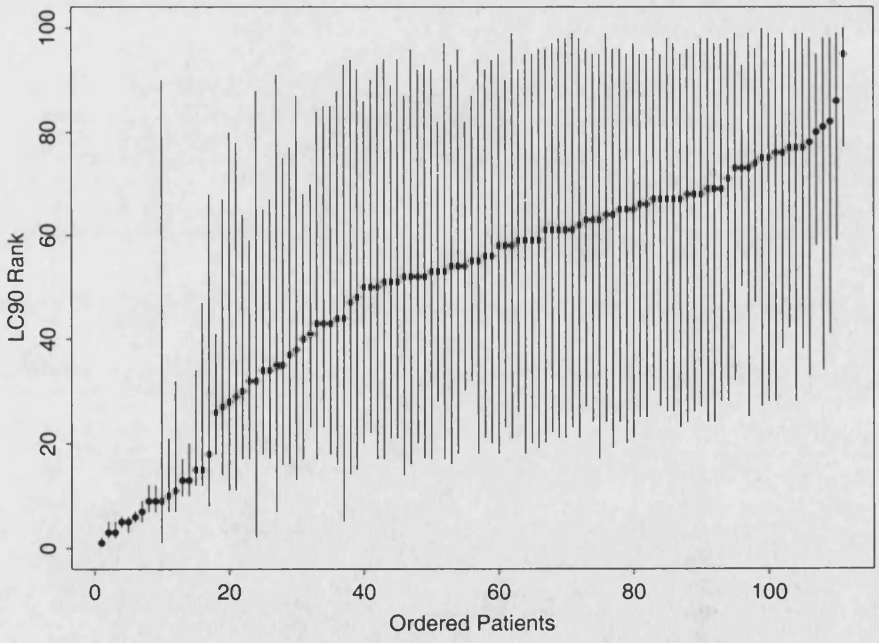

Figure 7-11: *Sorted LC90 Ranks for IFN with 5% and 95% percentiles.*

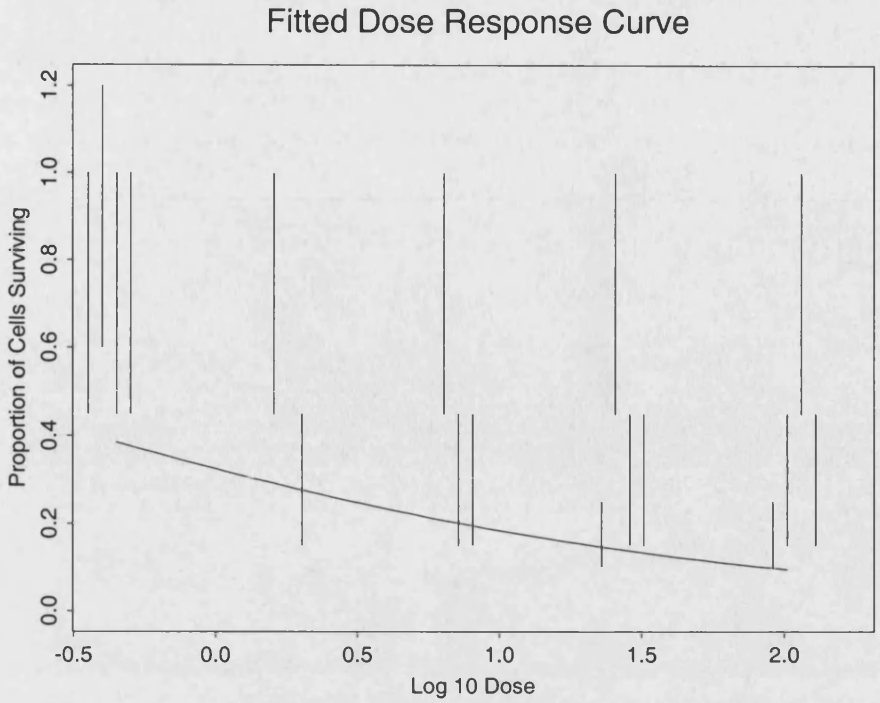

Figure 7-12: *Dose response curve for IFN patient 111.*

occasionally *increases* have been recorded which further distort the curve.

If the range of doses tested had been sufficiently broad to actually contain the LC90, the results would be more accurate. However, such high doses may not be tolerated by patients and more accurate estimates at such high levels could be impracticable to prescribe safely.

7.5 Individual patient results

We shall look at the results of individual patients in order to examine the variability of results within one subject.

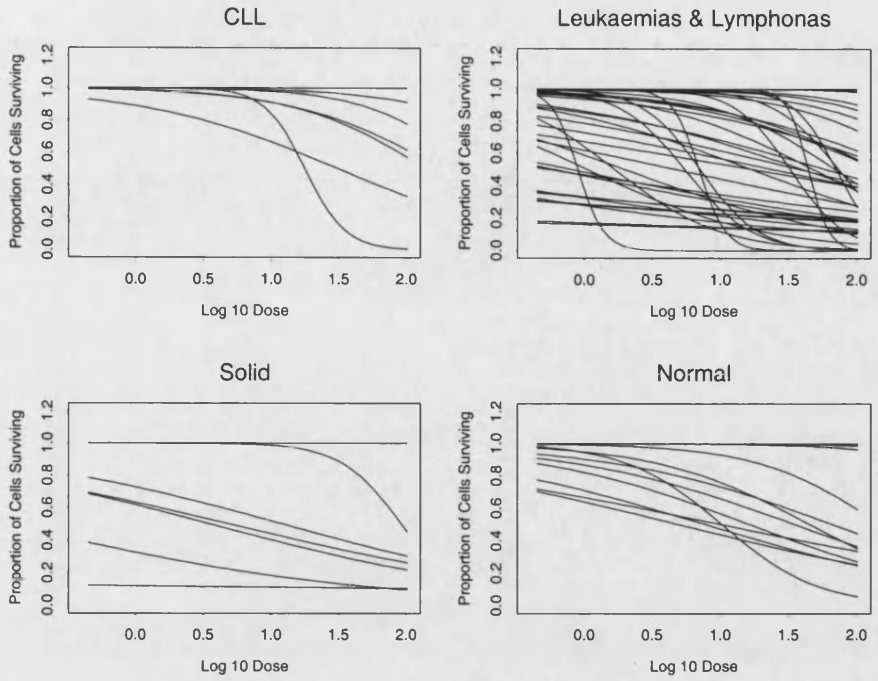

Figure 7-13: *IFN Dose response curves for all patients by tumour type.*

7.5.1 Patient number 1765

Table 7.3 shows all the assay results from patient number 1765. Of the 26 sets of results, there is a wide variety in slope between drugs which ranges from 0.5 to nearly 13. The $\log_{10}$LC90 also shows considerable variation.

The drug showing the "best" assay result in terms of the lowest estimated $\log_{10}$LC90 is MEP with a value of -1.252. The corresponding 5, 50 and 95 percentiles of ranks achieved are 0, 2, and 7. The next smallest estimated $\log_{10}$LC90 is CDA with -1.23: the corresponding ranks however are 40, 50, and 60. Thus the absolute values do not always agree well with their relative positions. This is to be expected as some drugs work well at lower doses and some require high doses to be effective.

The estimated dose response curves are shown in Figure 7-14. Whilst most drugs tested on this patient have curves with steep slopes, there are five drugs with very shallow curves which share similar characteristics. Two of these are the steroids

Drug	$\log_{10}$LC90	Percentile ranks 5%	50%	95%	b
AMS	-0.859	23	28	37	0.76
AR	-0.281	6	11	19	1.96
ASP	2.887	48	58	75	1.23
BC	1.076	7	11	17	6.70
BL	2.135	35	42	49	2.12
CB	1.743	49	59	69	7.55
CC	1.127	15	23	39	4.20
CDA	-1.232	40	50	60	2.64
CHL	0.401	28	33	39	4.20
DC	3.905	79	91	97	0.54
DN	-0.890	23	29	35	12.14
DOX	-0.96	12	15	22	7.21
EPI	-0.583	24	34	41	12.34
FL	-0.261	38	44	52	3.15
HIF	0.086	7	9	14	7.58
IDA	-0.987	37	43	47	10.21
LP	0.364	16	21	28	3.70
MAF	-0.032	12	19	27	12.81
MEP	-1.252	0	2	7	0.56
MP	2.60	21	27	39	9.27
MZN	-0.710	27	31	36	6.38
PR	-0.546	1	4	9	0.59
TG	1.929	55	64	72	3.38
VB	-1.06	1	4	9	0.82
VC	-1.082	3	8	15	0.63
VP	1.322	40	47	54	5.88

Table 7.3: *Assay results from one patient (No=1765).*

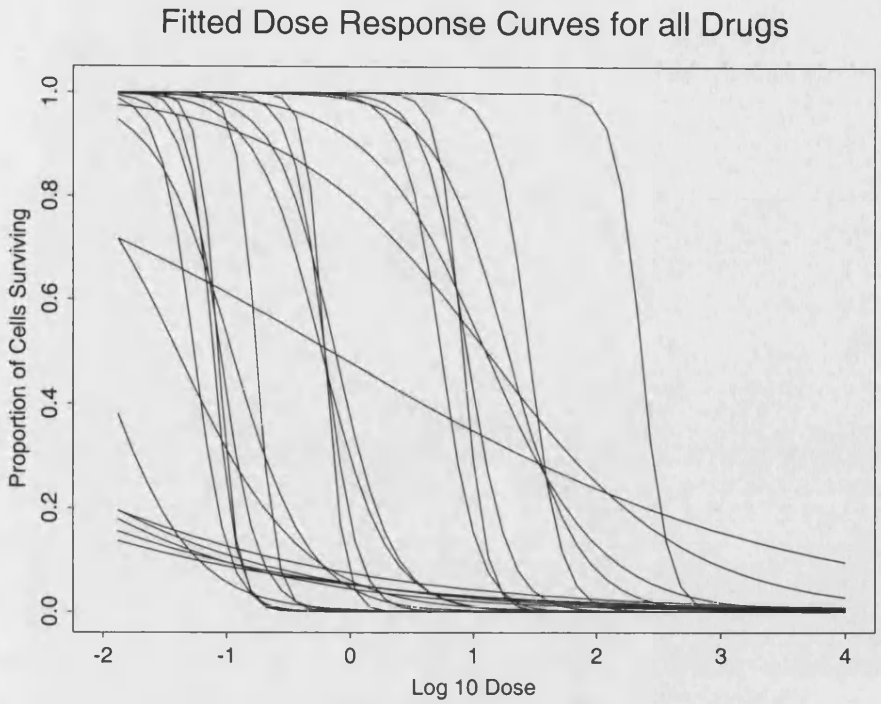

Figure 7-14: *Dose response curves for all assays performed for Patient 1765.*

tested (MEP and PR), two the vincal alkaloids used (VB and VC) and AMS is one of the two antibiotics examined (the other antibiotic, BL, has a steep curve). Both the vincal alkaloids and steroids have their median LC90 percentiles within the top 10% whereas AMS is ranked in the 28th percentile.

Of the drugs tested for this patient, there is a range in their relative responsiveness. Drugs with their median LC90 rank in the top 10% are HIF, MEP, PR, VB, and VC. Their posterior intervals are sufficiently narrow to provide confidence in these choices. There are also drugs showing a poor relative response, the least effective of these being DC which is in the bottom 10% of potential responders.

7.5.2 Patient number 1306

Table 7.4 shows the assay results from a patient with a poor response to nearly all the drugs tested. Again there is a wide variety in the estimated slopes of the

drug response curves, although these are much steeper than the previous example ranging from 2 to over 25. There is only one drug which indicates the patient may be in the top 50% of responders, FU, whereas all the others have poor predicted responses.

Figure 7-15 shows all the response curves for patient 1306, which have a similar shape over a range of locations. Ranking of this patient's LC90 with the whole population tested is highest for FU. The dose response curve for FU is not obviously any better than the other drugs tested in Figure 7-15 as it does not have a particularly small LC90 when compared on the same scale as the other drugs. However, the other results show lower percentile rankings so the decision to treat with FU will depend on whether the dose can be tolerated *in vivo*. Also the relative benefits of the other drugs tested may show lower estimated LC90s will depend on the overall efficacy of the drug, a lower percentile may not necessarily mean a worse response. It is especially important in cases like these that we are confident the type of assay used has a good clinical correlation i.e. a strong relationship between *ex vivo* sensitivity and actual response.

7.5.3 Patient number 1383

In contrast to patient 1306, patient 1383 has a very good response to the drugs tested. The figures are shown in Table 7.5. Although fewer drugs are tested 4 have their median values to be within the top 10%.

Figure 7-16 shows a variety of response curves, however they all show fairly steep curves and the LC90s have all been estimated within the doses tested.

Drug	$\log_{10}$LC90	Percentile ranks			b
		5%	50%	95%	
ACD	1.197	95	97	99	14.59
ACL	1.591	98	99	100	16.24
AMS	1.466	96	98	100	10.88
AR	3.042	87	92	98	7.26
ASP	2.905	41	55	88	6.40
BC	3.220	98	100	100	6.47
BL	2.727	61	66	72	6.70
BU	3.368	52	77	97	8.64
CB	2.133	84	94	98	24.62
CC	3.036	97	99	100	7.04
CDA	1.087	91	95	99	7.264
CHL	1.671	89	92	96	10.59
CP	1.587	97	98	99	7.51
DC	3.047	60	72	89	4.01
DEX	4.156	88	95	100	6.03
DN	0.393	96	97	98	9.44
DOX	0.487	90	93	94	11.26
EPI	0.425	92	95	96	6.86
FL	1.627	94	95	96	6.86
FU	2.053	34	40	48	6.00
GLA	3.557	20	52	92	10.13
HIF	2.928	99	100	100	8.47
HU	4.379	94	97	99	2.19
IDA	0.928	99	100	100	3.20
IFN	3.557	20	52	92	10.13
LP	1.924	95	97	99	10.42
MAF	2.461	95	99	100	14.36
MEP	2.835	81	92	98	13.70
MMC	0.711	77	87	93	2.69
MP	4.347	85	96	100	2.25
MZN	1.151	96	97	99	7.80
PR	3.726	99	99	100	25.68
TG	2.390	82	89	94	3.61
VB	1.920	86	90	96	9.61
VC	2.813	95	98	100	7.45
VD	2.420	72	89	98	9.32
VO	1.430	93	95	98	8.56
VP	3.146	93	97	100	8.06

Table 7.4: *Assay results from one patient (No=1306.)*

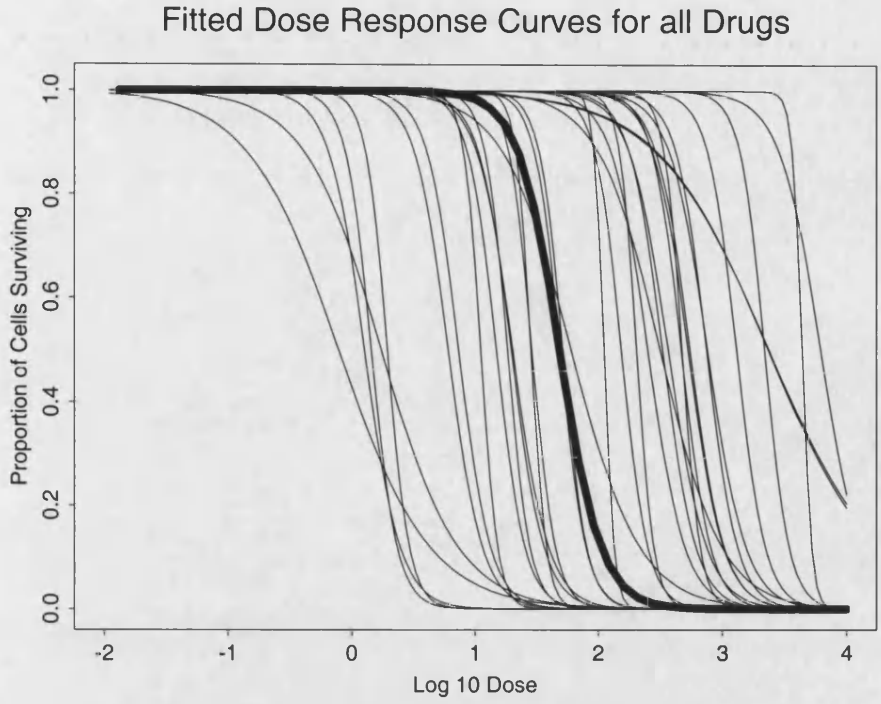

Figure 7-15: *Dose response curves for all assays performed for Patient 1306. The estimated curve for FU the drug with the highest percentile result is highlighted in bold.*

Drug	$\log_{10}$LC90	Percentile ranks 5%	50%	95%	b
BL	0.422	0	1	3	6.70
CC	1.199	17	32	47	7.04
CCA	0.994	8	17	37	7.04
CDA	-2.563	0	1	5	7.264
CHL	0.116	8	17	37	10.59
DOX	-0.951	9	16	27	11.26
EPI	-0.818	6	12	25	6.86
FL	-1.448	0	3	12	6.86
MAF	-0.296	3	7	14	14.36
MEP	1.792	47	55	63	13.70
PR	2.162	41	50	61	2.69
VC	-0.288	18	23	30	2.25

Table 7.5: *Assay results from one patient (No=1383).*

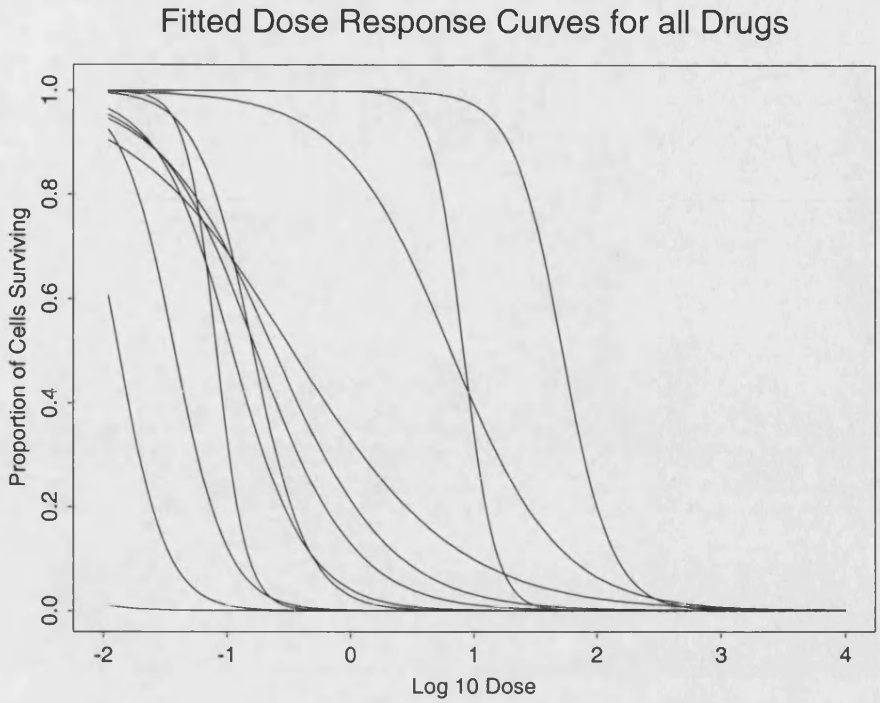

Figure 7-16: *Dose response curves for all assays performed for Patient 1383.*

7.6 Future subjects

The results reported so far are from the analysis of complete sets of data gathered over a number of years. When further assays are performed on a new subject, we could add these into the existing data sets and run the full analyses again for each drug tested, processing data for previous patients again as well as that for the new patient. The main disadvantages in doing this are:

- A big calculation is involved again, a lot of which is unnecessary.
- The people performing the assays and collecting the data at the hospital would prefer to have a smaller program which is quicker to run and requires less maintenance.

It is a feature of Bayesian updating that data can be processed in batches and population parameter distributions updated with each new set of data. The posterior distributions from one analysis serve as the prior for the next. Consider

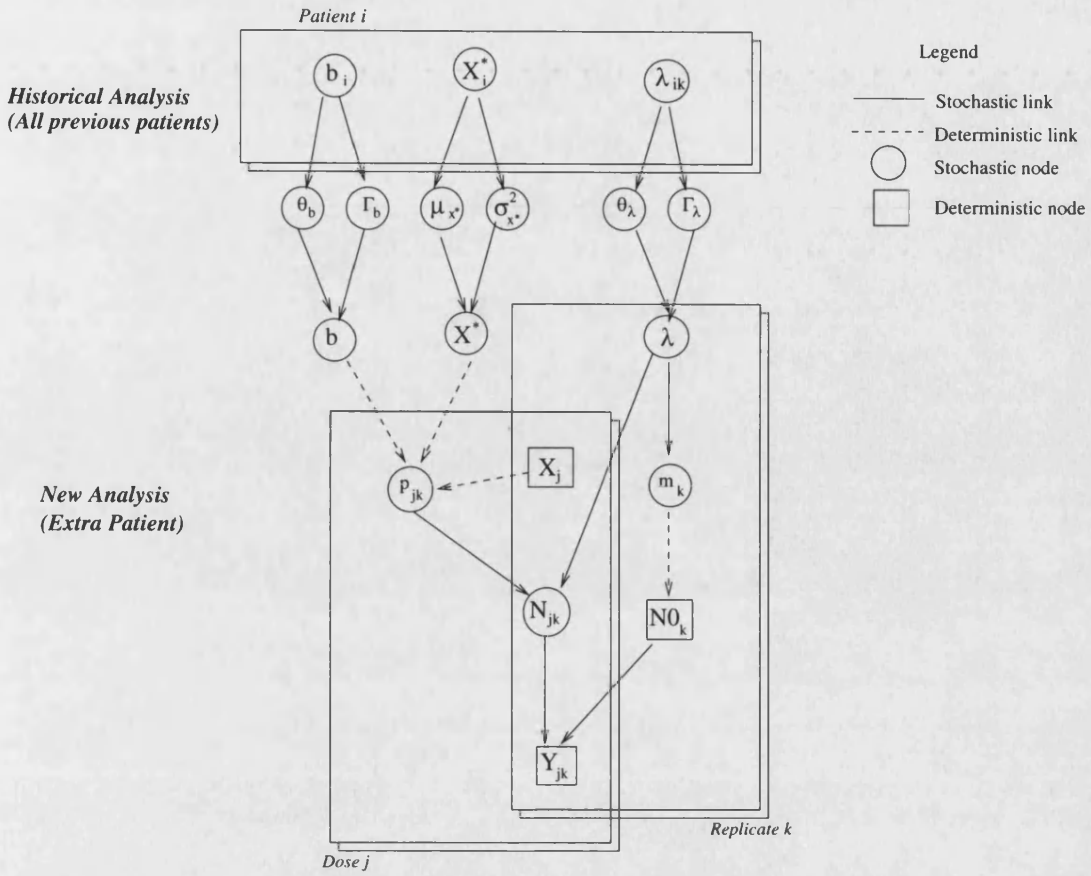

Figure 7-17: *Directed Acyclic Graph of model for extra subjects.*

the particular problem of analysing assay results for a new case after a large historical set of data. In order to analyse the new case, it is sufficient to know the posterior joint distribution of λ, b, and x^*: the initial cell density, slope and LC90 respectively, for a new subject. We have therefore obtained these posterior joint distributions for each drug and cell type combination.

7.6.1 New model parameters

Figure 7-17 shows the existing model modified to accept information about posterior distributions from the initial analyses which is in turn used as prior information for the less complicated model required for just one extra case. The *historical* analysis in this figure only shows the important variables from which

information is utilised in the second model. For the complete diagram of the first model, see Figure 4-6.

From our initial analysis, the instinctive approach would be to build up a posterior distribution of $\theta_\lambda, \kappa_\lambda, \theta_b, \kappa_b, \mu_{x^*}$ and $\sigma^2_{x^*}$, and use this to define a sampling distribution $y(\theta_\lambda, \theta_b, \mu_{x^*})$.

Another technique to obtain the sampling distribution would be to include a "new patient" as an extra case in the model and keep updating $y(\theta_\lambda, \theta_b, \mu_{x^*})$ for this case and then model its posterior distribution. However, this method has smaller variation in the posterior distributions for the variables we are interested in since it is considering just one "typical" patient. The between patient variation is considerably larger than the within patient variation for all variables and thus we would underestimate the overall expected variation.

The method employed has utilised the posterior distributions obtained from λ, b, and x^*. We can fit standard distributions to these variables by deriving estimates for the distribution parameters from the relevant posterior distributions. We now use these in place of the original hyperpriors ($\theta_\lambda, \kappa_\lambda, \theta_b, \kappa_b, \mu_{x^*}$ and $\sigma^2_{x^*}$).

This method has the advantage of not being tied to the same distribution as specified in the initial prior model where we specified diffuse prior distributions.

7.6.2 Deriving prior distributions

In order to use the posterior distributions obtained from the initial analyses as prior distributions for analysing new subjects, we must find some way of fitting suitable distributions which describe them adequately. Firstly we need to check whether the variables can be considered as independent.

We shall take the drug ACD as an example. To test independence, we can draw scatter plots as in Figure 7-18. The points represent the posterior means for the main variables λ, b, and x^* taken from all patients regardless of sample type. Using these graphs we can perform visual checks for whether there are any discernible trends: the plots in this case all look unrelated. A further check is to calculate correlation coefficients, as in Table 7.6. These values are all low and

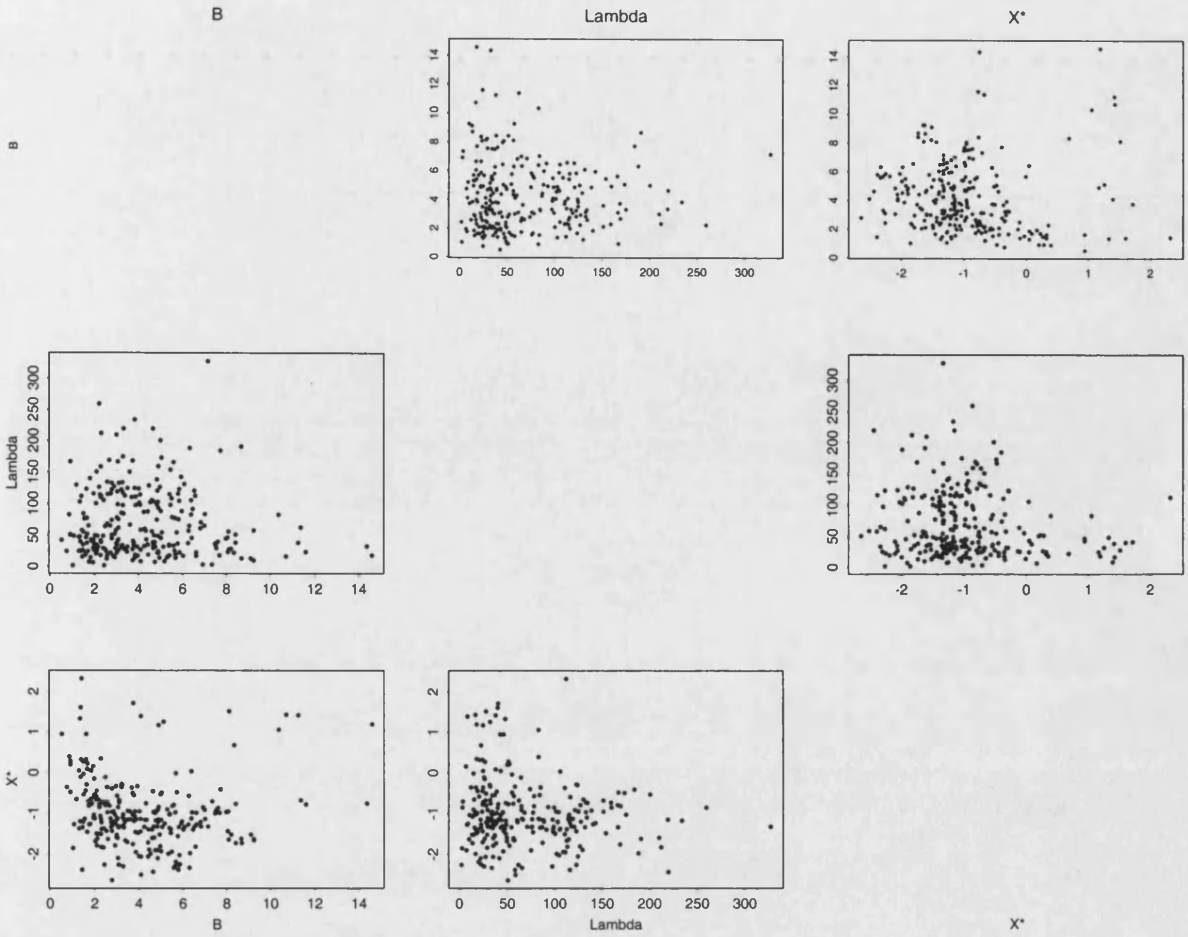

Figure 7-18: *Matrix of scatter plots for ACD.*

show either negligible correlation or are sufficiently small to cause no concern. Hence there is no evidence of any demonstrable relationship between these variables and so they can be considered independent for subsequent purposes.

We can now model the distribution of λ, b, and x^* which we shall use as prior distributions for the parameters of a new subject. Since both λ and b are required to be positive, as our starting point we use Gamma(θ, κ) distribution, with mean$=\theta\kappa$ and variance$=\theta^2\kappa$.

As x^* is on the $\log_{10}$ Dose scale, negative values are permitted and so a Normal distribution can be used as a prior distribution taking the posterior mean and variance as its parameters.

	B	λ	X^*
B	1	-0.053	0.143
λ	-0.53	1	-0.043
X^*	0.143	-0.043	1

Table 7.6: *Correlation matrix for the posterior distributions, Drug=ACD.*

A table has been constructed which contains all the posterior means and standard deviations of b, λ and x^* for all drug and tumour cell combinations so that new prior distribution parameters can be evaluated. See for example Table 7.2; the values vary considerably between disease type for each drug as well as across drugs. Posterior distribution summaries of the initial hyper-prior variables are not given since new hyper-prior variables have new parameters calculated from the posterior distributions of their offspring.

To check that the distributions chosen for the priors are appropriate, we can use graphical methods. To assess the goodness of fit, we can overlay the theoretical density functions on the posterior distribution histograms. For example we look at the drug ACD. Figure 7-19 shows the fitted distribution as a dotted line. In all cases the fits seem satisfactory. A further check is to use qq plots, as in Figure 7-20. A perfect fit would show all points to lie on the diagonal. Any systematic deviation from this indicates a poor fit: slight departures, especially at the extremities are to be expected. Other drugs show broadly similar results. However since these fitted distributions are used to provide prior information, the subsequent analyses are fairly robust to the actual prior distribution used, it is not necessary to obtain perfect fits.

Whilst it may initially seem that shrinkage may occur in deriving new parameter estimates from λ, b, and x^* rather than simply using the posterior distributions of the θ_λ, θ_b, μ_{x^*} and $\sigma^2_{x^*}$, this is not actually the case. The estimated variance for x^* from the posterior of this variable is either larger or of a very similar magnitude when compared to the estimate derived from the posterior distribution of $\sigma^2_{x^*}$. The estimated variances of both λ and b are consistently larger when using the relevant posterior distributions to derive these values rather than those from the respective hyperprior posterior distribution.

This occurs because the inter-patient variation is considerably larger than the

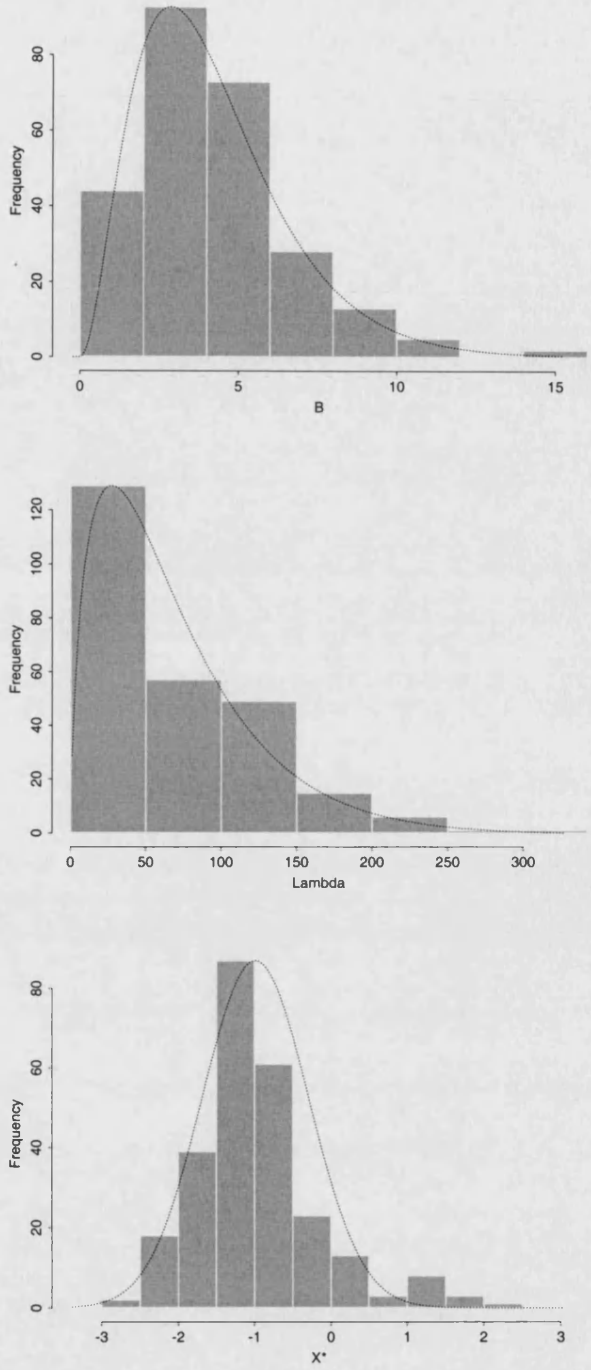

Figure 7-19: *Histograms and fitted distribution for drug=ACD.*

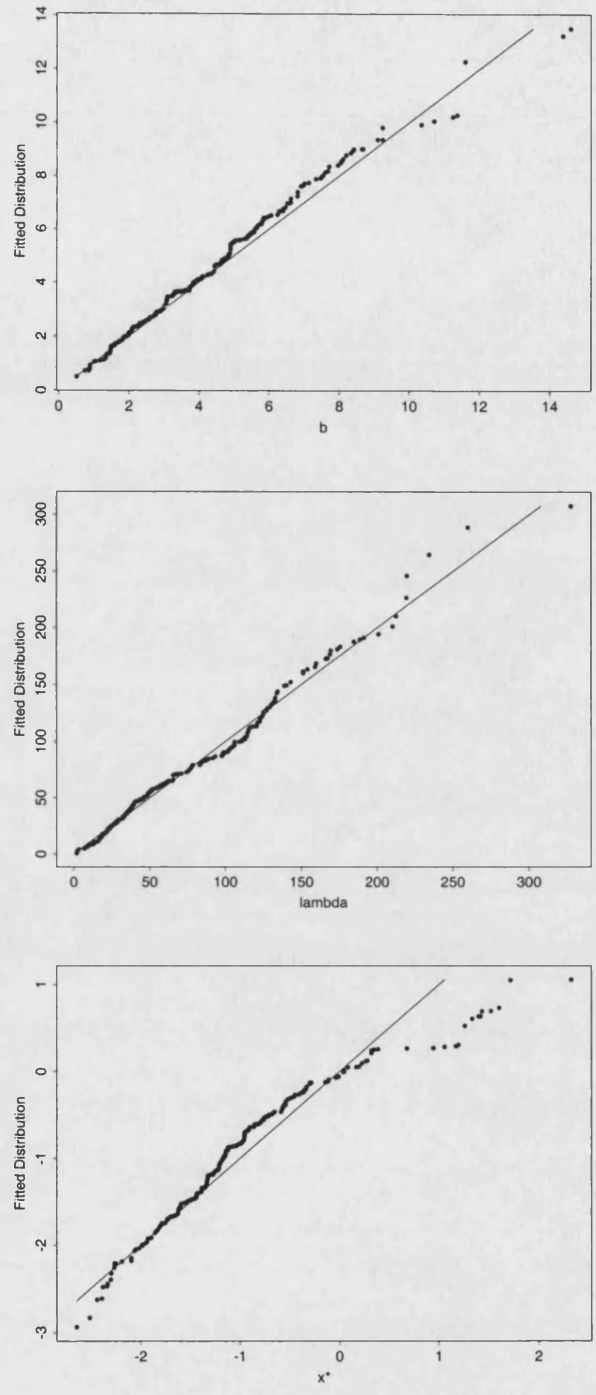

Figure 7-20: *QQ plots for drug ACD.*

within patient variation. Using information which takes into account the posterior distributions of all patients rather than merely a distribution parameter of this posterior distribution gives a similar or bigger variation.

In order to estimate any future subjects, we require the following information: the control values, the proportions of surviving cells and the dose levels used. Together with the drug and tumour type, the parameters for the new prior distributions can be estimated.

7.6.3 Analysing a new case

We define $\pi_0(\lambda, b, x^*)$ as the posterior distribution of λ, b, x^* for one new subject following the analysis of the complete historical data. The new posterior distribution of (λ, b, x^*) given y can now be evaluated:

$$\pi(\lambda, b, x^*) = \frac{\pi_0(\lambda, b, x^*) l(y; \lambda, b, x^*)}{\int_{\lambda'} \int_{b'} \int_{x^{*'}} \pi_0(\lambda', b', x^{*'}) l(y; \lambda', b', x^{*'}) \mathrm{d}\lambda' \mathrm{d}b' \mathrm{d}x^{*'}} \tag{7.1}$$

where π_0 is the new prior distribution derived from the previous posterior distribution and $l(.)$ is the likelihood of the observed data.

The integral can be evaluated numerically using, for example, Simpson's rule. If the problem had been of a higher dimension, this approach would not perhaps be possible and running further models using the MCMC approach adopted previously may have been more appropriate.

A program has been provided to the RUH by Jennison which uses the posterior distributions of (λ, b, x^*) obtained from the main analyses and estimates these parameters for the additional case by evaluating Equation 7.1 above.

This new method of analysing additional cases has allowed us to investigate the method of data collection and to propose new strategies. In the historical data analyses, categories were of varying widths depending on whether the surviving proportions were large or small. They also overlapped by often considerable margins. There is scope for making precise counts of cells when there is a low underlying count (control value) or when numbers of surviving cells are small.

We have found that it is relatively efficient to "guesstimate" when the cell survival rate is high but exact counts are desirable otherwise since this reduces the scope for error. In consultation with the scientists at the RUH involved in the performing the assays and evaluating the results, this has led to a new protocol in cell counting. Obviously time is a factor, so exact counts are only made when numbers are sufficiently low for this to still be an efficient way of determining survival proportions.

7.6.4 Updating posterior distributions

It would be nice to keep on updating the posterior joint distribution of (λ, b, x^*) as more subjects are analysed. However, this is not so easy to implement and not too critical if there is sufficient historical data so the posterior distribution of (λ, b, x^*) really does reflect the population distribution accurately, i.e. with little extra variability due to uncertainty about the actual population distribution.

We can check the future results are consistent with the historical sample by looking at the combined posterior distribution of all the new cases. A more time consuming way would be to run the MCMC of the initial model with *all* the new data added and see how this affects the posterior distributions first obtained.

Without sufficient data this is impossible to check satisfactorily but performing analyses on just part of the data show a good agreement where there is sufficient data to start. With sufficient further data, we can investigate how large an initial sample is required before there is no appreciable increase in accuracy of the estimates.

7.7 Conclusions

The value of performing these assays and obtaining accurate estimates of the LC90 is demonstrated by showing the variability of cell survival curves both between drugs and sample types. It is impossible to predict in advance whether a patient will respond well to some drugs and not others. The DiSC assay used for

these studies shows a clinical correlation of 84% according to Bird *et al.* (1988) indicating that it is a good prognostic indicator of actual clinical response as well as a means of providing information on the sensitivity of a tumour to specific drugs. We can thus be confident that the results obtained here provide a very useful basis for prescribing treatment.

Analysing all the historical data means we are able to use the posterior distributions obtained as prior distributions when analysing new cases. This gives a real benefit in terms of efficiency as there is a smaller model to consider and a minimal amount of data.

Chapter 8

Outliers

We have already seen that the recorded categories for the proportions of surviving cells often differ between replicate slides or between counters. This may be due either to the actual differences in the numbers of cells between the two sets of slides or to the different assessment of live and dead cells between the scientists. When the slides are stained, the dye makes the live cells appear magenta under the microscope whereas the dead cells show up black. However, there are many dying cells which fall between these colours and it is often not clear cut how these should be categorised. This choice is therefore necessarily subjective and can lead to substantially different assessments of the proportions of surviving cells especially when the cells are sparse. We wish to have some idea of how these discrepancies may affect the results obtained and whether such discrepancies have much effect on the estimated LC90s.

8.1 Outlier definition

Barnett and Lewis (1978) define an outlier to be an observation (or subset of observations) which appears to be inconsistent with the remainder of that set of data. Outliers in this problem include both those which initially appear discordant with the rest of the data values and values which subsequently deviate substantially from fitted model.

Young and Pettit (1996) consider ways of measuring discordancy between the data and the assumed prior in order to test for the presence of outliers and also the adequacy of the model. After seeing the data they use Bayes factors to check whether the prior elicited from an expert (in our case, the scientists at the RUH) is misguided and that the data do not conform to their expectations. However, they concentrate on discordancy between the prior and data using Bayes factors, we are more interested in how our *posterior* fits the data.

There are two basic mechanisms which give rise to apparent outliers (Hawkins, 1980). The first arises when the data come from some heavy tailed distribution such as a t-distribution. The second involves two distributions: a basic distribution which generates well behaved observations and a "contaminating" distribution which generates "contaminants". The tendency is for extreme observations to come from the latter distribution.

8.2 Bayesian approach to outliers

If a model is to be believed, the posterior distribution allows all relevant estimation inferences to be made about its parameters (Box, 1980). Thus, while individual values may appear inconsistent with the model, these are not necessarily of importance and we look at the influence these values may have.

Barnett and Lewis (1978) outline some Bayesian methods to deal with outliers. Since such techniques require an *a priori* statement about the possible values of parameters in a parametric family of models, this must include a prior assessment of probabilities attached to the presence of outliers. Therefore we must anticipate the possible presence of outliers and structure our data generating model accordingly before analysis of any data.

It is generally agreed that it is desirable to accommodate possible outliers regardless of the mechanism assumed to give rise to them e.g. Besag & Higdon (1999), Barnett & Lewis (1978). Just as there are two mechanisms from which we assume outliers arise, there are two main ways of dealing with them. The first is to ensure the prior distribution supports the possible occurrence of outliers by

using non-informative or even improper priors, see, for example Young & Pettit (1996), the second assumes outliers arise from distributions other than that of the specified prior, eg Kale & Sinha (1971).

Kale and Sinha (1971) propose a model in which $n-k$ of the observations $x_1, x_2, ..., x_n$ arise from some basic population F whilst the remaining k observations (the outliers) arise from populations $G_1, G_2, ..., G_k$ distinct from F. It is assumed that before recording the observations, there is no way of identifying the anomalous subset or its size, k. It is also assumed that any subset of size k of the n observations is equally likely to be the set of observations arising from $G_1, G_2, ..., G_k$.

This is termed the *exchangeable model* since we assume that joint distribution of the random variables $(X_1, X_2, ..., X_n)$ is the same as that of $(X_{j_1}, X_{j_2}, ..., X_{j_3})$ for all permutations $(j_1, j_2, ..., j_n)$ of $(1, 2, ..., n)$.

This is in keeping with de Finetti's (1961) theory of exchangeability and his argument that any approach to outliers be couched in such terms.

This type of approach requires alternative distributions to be specified yet this is difficult to do since often we do not know an appropriate alternative since we have little information about the outliers. Young & Pettit (1996) use Bayes factors to compare the discordancy of different prior models and can then proceed to select the models with the least discordant priors.

Box (1980) states that each model under consideration can be imagined as being embedded in a more complex one. The model *is* the prior in the wide sense that it is a probability statement of all the assumptions currently to be tentatively entertained a priori. We write our model as

$$\mathrm{P}(y_D, \theta|A) = \mathrm{P}(\theta|y_D, A)\mathrm{P}(y_D|A)$$

where y_D is the observation being examined, θ is the vector parameters, and A represents the conditionality assumptions made in the model.

We can partition y_D into y^* and $y_{D'}$ where y^* is the outlier in question and $y_{D'}$

is the rest of the data so we now have:

$$\mathrm{P}(y^*, y_{D'}, \theta|A) = \mathrm{P}(\theta|y^*, y_{D'}, A)\mathrm{P}(y^*|y_{D'}, A)\mathrm{P}(y_{D'}|A).$$

The second factor:

$$\mathrm{P}(y^*|y_{D'}, A) = \int \mathrm{P}(y^*|y_{D'}, \theta, A)\mathrm{P}(\theta|A)d\theta$$

is the predictive density associated with the value y^* actually obtained. If the y^* was unlikely to be generated by the model, this could be assessed by reference to the density $\mathrm{P}(Y|A)$ or of the density $\{g_i(y^*|A)\}$ of some relevant checking function $g_i(y^*)$. In making this check, it is not necessary to be specific about an alternative model.

Gelfand *et al.* (1992) consider outliers and residuals as way of assessing model adequacy. They note Box (1980) states that the posterior distribution provides a basis for *estimation* of parameters conditional on the adequacy of the model whilst the predictive distribution enables *criticism* of the model in the light of the current data. They proceed to explain that in model comparison predictive distributions are directly comparable whilst posteriors are not. The predictive distribution (or marginal likelihood) is the joint marginal of the data and may be used for model determination.

The predictive distributions $f(Y_r|y_{(r)})$ are to be checked against y_r for $r = 1, 2, ...n$ so that if the model holds, y_r may be viewed as a random observation from $f(Y_r|y_{(r)})$, where Y denotes a $n \times 1$ data vector, y_r is the r^{th} value from our sample with $y_{(r)}$ the $(n-1) \times 1$ vector with observation r omitted. To assess the model adequacy, Box (1980) advocates the use of a checking function, $g(Y_r|y_r)$, whose expectation under $f(Y_r|y_r)$ is calculated and denoted by d_r. The strategy employed is a Bayesian analogue to the frequentist version of examining studentised residuals. Several choices for g are available to us but all have the characteristic that many large $|d_r|$ cast doubt upon the model.

Standard Bayesian methods for model *choice* given J proposed models M_j denoted as $f(Y|\theta_j; X, M_J)\pi(\theta_i)$ also makes use of the predictive distribution.

Let w_j denote the prior probability of M_j. Using Bayes Theorem, the posterior probability of M_j is:

$$p(M_j|Y) = f(Y|M_j)w_j / \sum_{j=1}^{J} f(Y|M_j)w_j \tag{8.1}$$

where $f(Y|M_j)$ is the predictive or joint marginal distribution of the data under model M_j. For observed data, y, the model yielding the largest $p(M_j|y)$ is selected.

Gelfand *et al.* (1992) suggest this approach is impractical and unrealistic and proffer a maximum expected utility approach which modifies this basic idea. An alternative form is to choose M_1 if :

$$E_{f^*}\left[\log \frac{\Pi f(Y_r|y_{(r)}, M_1)}{\Pi f(Y_r|y_{(r)}, M_2)}\right] > 0 \tag{8.2}$$

where $f^* = w_1 \Pi f(Y_r|y_{(r)}, M_1) + w_2 \Pi f(Y_r|y_{(r)}, M_2)$. If this condition is not satisfied, model M_2 is chosen.

This expected utility approach is readily extended to $J > 2$ models.

The approach of Besag and Higdon (1999) towards the assessment of outliers involves substituting the Gaussian distribution usually assumed for residuals by a contaminated Gaussian one with heavier tails such as the Student's t_{ν_y} where ν_y itself has a prior distribution. They call this a *hierarchical* t-formulation with the prior for ν_y to be uniform on (1, 2, 4, 8, 16, 32, 64) although of course any other choice could be made.

Realistically, model criticism is often conducted by visual inspection of residual displays designed to highlight "features" in the data that would rarely be extreme if the model were true. If such a feature can be described by a function $g_i(y^*)$, its unusualness, if formalised, would be measured approximately by reference to $\mathrm{P}\{g_i(y^*|A)\}$.

The distinction between parametric features of the model and residual features is somewhat arbitrary. In practice the needs of parsimony urge us to settle for

reasonably simple models and to consider possible deviations from them.

Of the two main types of approaches outlined for dealing with outliers, the first, requiring specification of alternative prior distributions from which outliers are said to belong, is too cumbersome to be used in this particular problem. The approach of Box (1980) is more appropriate in that no alternative distributions are required. This is especially relevant given the non-standard distributions with which we are mainly concerned.

8.3 Outlier detection

Detection of outliers can be very subjective if we do not know *a priori* the likely patterns of outlying values of the variables. In this problem we are concerned that several of the y-values do not "fit" with the rest of the observations and we wish to assess the impact of these atypical values. Such data can be referred back to the scientists who counted the cells, who can re-examine the slide and possibly provide new counts, or, if the slide is found to be of poor quality, simply remove the relevant data points.

8.3.1 Initial screening

In order to detect possible outliers we must first devise some algorithm for identifying them automatically. Given the size and nature of the data set being examined it would be impractical any other way. It is also less subjective and far more efficient since algorithms can be incorporated into the existing programs.

Figure 8-1 shows the different probabilities of being in each category according to the actual proportion of surviving cells. Most values may fall in one of two categories since these categories overlap, often substantially. The lower categories are much narrower where more care is taken to ensure accurate counts are taken when cell survival is low.

Thus we define observations which differ by more than one category to be *discordant.* If a recorded observation is observed to be discordant with all the

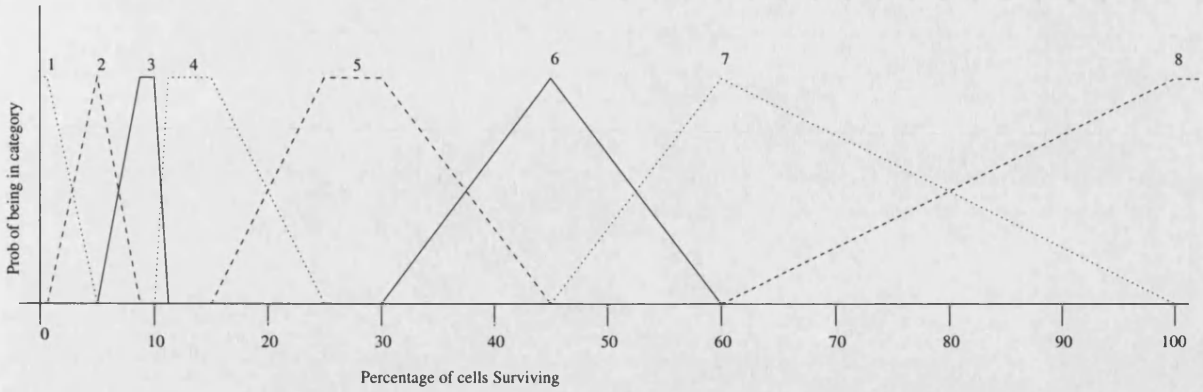

Figure 8-1: *Probability of being in recorded category given proportion of surviving cell.*

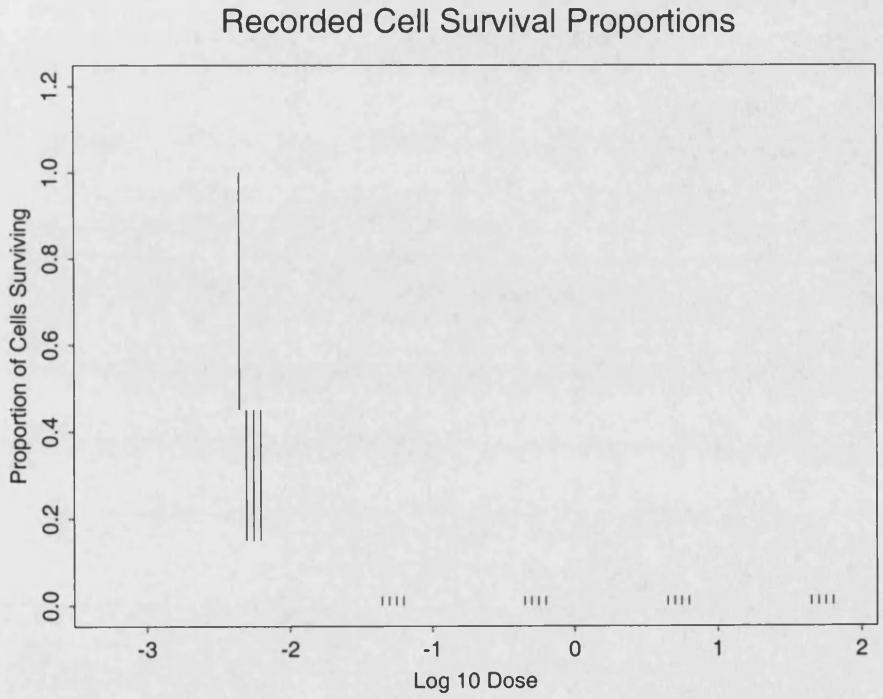

Figure 8-2: *Drug ACD, Patient 4. Proportions of surviving cells.*

other y values we flag it as a possible outlier. For example, see Figure 8-2; whilst the 4 higher doses all agree, the smallest dose has three values of Y at category 5 (15–45%), the final replicate has a recorded value of 7 (45–100%). If an observation is found to be discordant, we take this as an indication it should be analysed more closely with a view to declaring it as "outlier" and either removing it from the analysis or referring it back to the scientists for re-assessment.

Obviously we need more than 2 sets of replicates to determine this as otherwise it is impossible to say which of two observations is the more unlikely without further data being available. Therefore only subjects with at least three replicates are considered in this manner.

Another type of outlier in this type of data may occur when the recorded categories are decreasing non-monotonically and the value does not appear to fit with the rest of the data. An example of this is shown in Figure 8-3. In this case it is harder to detect which value is likely to be an outlier, the fourth or fifth value as this would depend on the slope of the dose response curve. If further replicates were available, then this may help with the identification of the erroneous value.

8.3.2 Formal assessment of outliers

In assessing a value flagged as a possible outlier we exclude the observed value from our analysis just as we would treat a missing data value. In the MCMC analysis, new values are proposed for the "missing" observation and accepted or rejected according to the probabilities given in Section 5.6.2. In this way we can look at the resulting distribution of values for this observation generated in our Markov chain and compare this distribution with the value actually recorded. We then have some information of how inconsistent the recorded value is when compared to predictions based on the rest of the data. If the recorded value is inconsistent with these predictions we declare the observation as an outlier. This approach departs from a strict application of de Finetti's model as we have not specified just how we expect observations to behave when they do not follow the standard model: although such a specification might be possible in priniciple, it

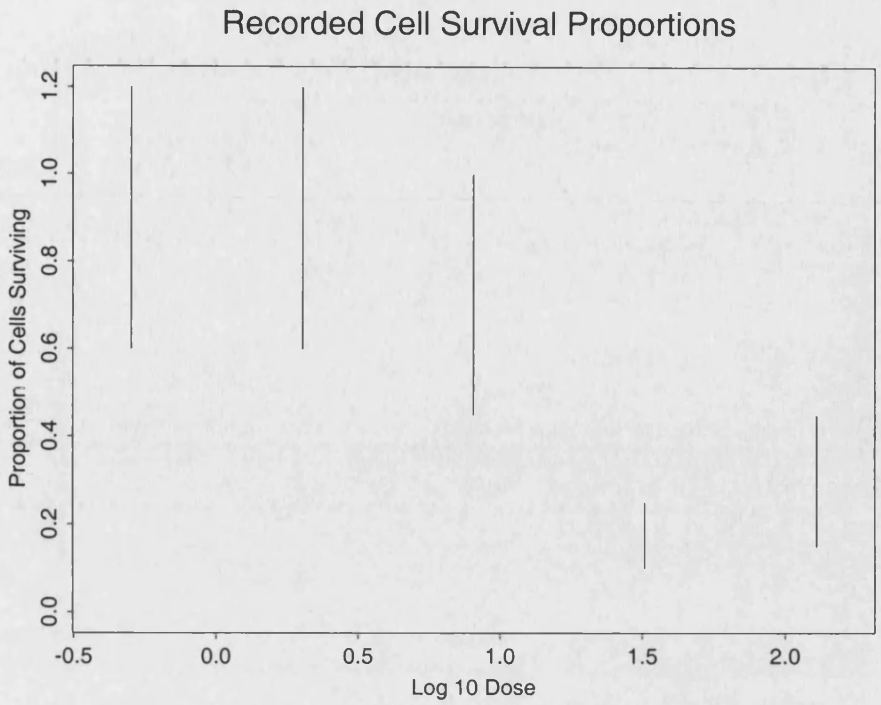

Figure 8-3: *Drug IFN, Patient 48. Proportions of surviving cells.*

is not clear how one should start to create one and we are happy to sidestep this issue in our treatment (Box, 1980).

8.4 Results

We shall use the drug ACD to illustrate the methods and typical results achieved. Of the 258 patients with ACD assays, 47 had sufficiently discrepant values recorded to result in possible outliers being flagged. These values were treated as missing and the model as described in Chapter 4 was run again so we can compare the sets of results.

The drug ACD is "well behaved" in that the recorded survival proportions (Y_{ijk}) are all non-increasing with increasing dose in all but one assay (Patient 32). As mentioned in Section 8.3.1 the second type of outlier is harder to detect as it depends on the neighbouring observations and the underlying model. We shall use the drug IFN as an example of an "ill behaved" drug with survival proportions

		Dose: $\log_{10} \mu g/\text{ml}$	0.0005	0.005	0.05	0.5	5
Replicate	1	Slide 1 Counter 1	5	1	1	1	1
	2	Counter 2	5	1	1	1	1
	3	Slide 2 Counter 1	5	1	1	1	1
	4	Counter 2	7	1	1	1	1

Table 8.1: *Drug ACD, patient 4: Recorded Y values (categories for proportion of surviving cells).*

actually increasing with drug dose. The presence of discrepant values does not generally alter the LC90 estimates. However, if a marked proportion *are* affected significantly, this also affects the associated rankings which may change by several percentile points.

The first record we encounter with a potential outlier for the drug ACD is patient 4.

ACD, Patient 4

In this case, the value flagged as discrepant is replicate 4, at the lowest dose level. The recorded category is 7 (45% - 100%) as opposed to the other three recorded values of 5 (15%–45%), see Table 8.1. The resulting Markov Chain when treating this value as missing yields the posterior distribution as shown in Figure 8-4.

The posterior mode is 5 which agrees with the other values recorded. This is achieved 46.3% of the time whereas the actual value recorded, 7, is visited only 1.7%.

Table 8.2 shows the estimated parameter values for both using the value as recorded and treating it as missing.

The estimates for X^*, the $\log_{10}$LC90 do not alter much, nor do the associated rankings. However the slope, B, is much lower with a much reduced standard deviation indicating a more reliable estimate, see Figure 8-5 to see how this affects the estimated dose response curve. This is to be expected since the estimated Y value is now the same as the other replicates for nearly 50% of the updates

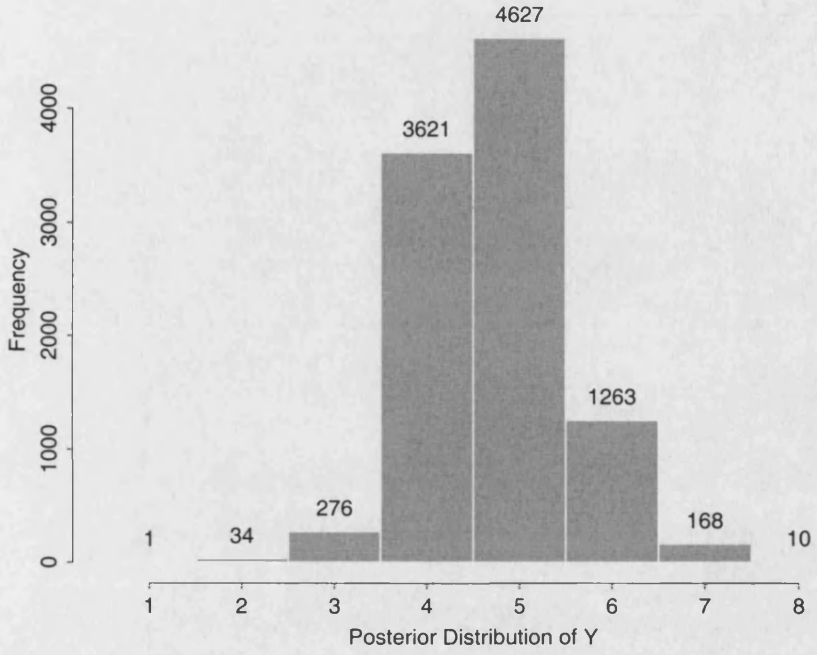

Figure 8-4: *Drug ACD:Posterior distribution for* $Y_{4,4,1}$.

					Ranks		
	E(X^*)	sdX^*)	E(B)	sd(B)	5	50	95
Initial Estimate	-1.93	0.113	5.01	1.49	7	10	16
New Estimate	-1.98	0.103	3.59	0.78	6	9	13

Table 8.2: *Drug ACD, Patient 4. Parameter estimates both before and after omission of potential outlier.*

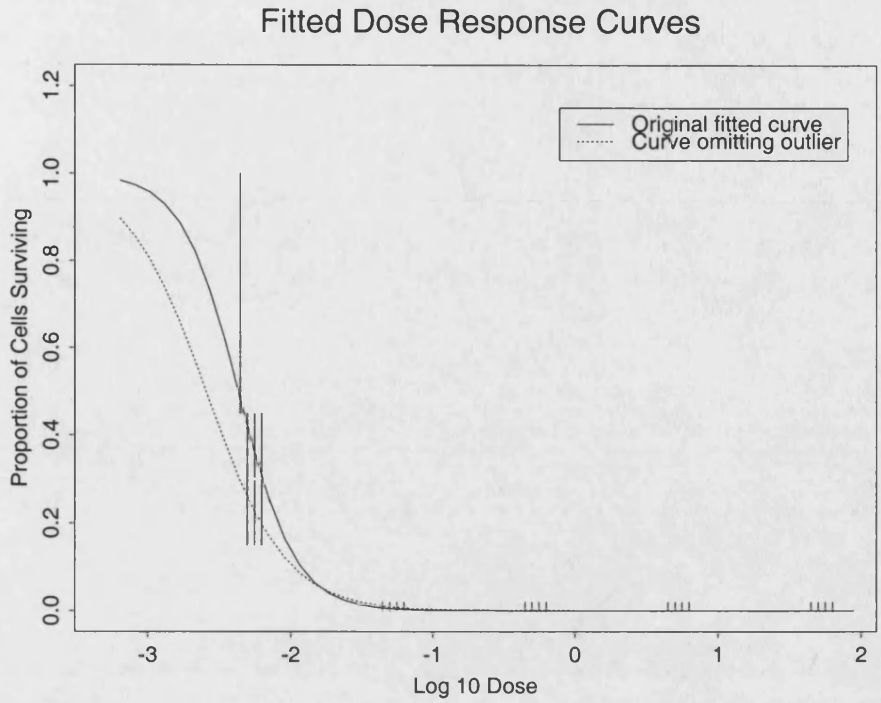

Figure 8-5: *Drug ACD, Patient 4. Proportions of surviving cells.*

performed. It is worth noting that the control value, the underlying cell density for an untreated slide, for the patient is 113 and any changes in the number of surviving cells will not drastically alter the survival proportions and consequently the LC90s. This particular example shows little difference in the LC90 and so, although the recorded value may be regarded as atypical, being visited less than 2% of the time, there is little overall effect to concern ourselves about.

We shall look at further examples of where the control value is small and this does have a bearing on the estimated LC90.

ACD, Patient 48

The next example we shall consider is patient 48, also the ACD assay. Table 8.3 shows the recorded values. The possible outlier here is replicate 1, dose level 3. The control value recorded for this assay is 36.3, which is relatively small.

Figure 8-6 shows the posterior distribution achieved after running the model. It

		Dose: $\log_{10} \mu g$/ml	0.0005	0.005	0.05	0.5	5
Replicate	1	Slide 1 Counter 1	8	7	7	1	1
	2	Counter 2	8	7	4	1	1
	3	Slide 2 Counter 1	8	7	5	1	1
	4	Counter 2	8	7	4	1	1

Table 8.3: *Drug ACD, patient 48: Recorded Y values (categories for proportion of surviving cells).*

					Ranks		
	E(X^*)	sdX^*)	E(B)	sd(B)	5	50	95
Initial Estimate	-0.98	0.126	2.29	0.40	62	72	79
New Estimate	-1.13	0.114	2.80	0.53	36	49	60

Table 8.4: *Drug ACD, Patient 48. Parameter estimates both before and after omission of potential outlier.*

can be seen that the recorded value was visited 0.003% of the time and the mode was 4, with 54.3% of the time spent at this value. This agrees well with the other recorded values.

The corresponding parameter estimates are given in Table 8.4. In this case, treating the potential outlier as missing does have an affect on the LC90. Bringing this atypical value to resemble the other replicates reduces the LC90 somewhat from -0.98 to -1.13 on the $\log_{10}$ scale. The corresponding ranks change considerably so they do not overlap. The initial median rank is 72, and the new one is 49, although this does have a slightly bigger range.

As previously stated, this patient has a relatively small control value so small differences in assessing the cell survival rates do greatly affect the LC90, as can been seen in this case.

IFN, Patient 48

This example shows a case where the proportions of surviving cells show an increase with dose, the last recorded value being greater than the previous observation, see Figure 8-3. As stated in Section 8.3, without some indication of

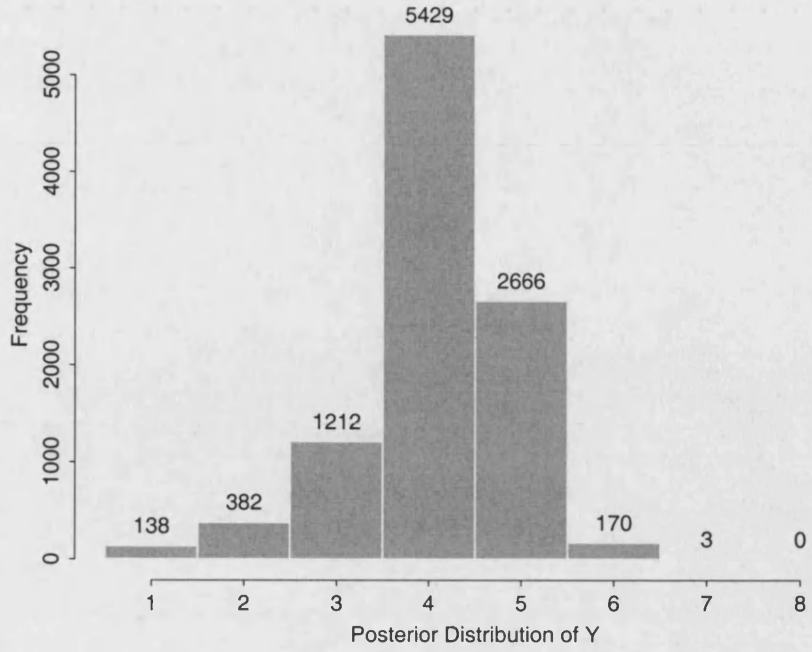

Figure 8-6: *Drug ACD:Posterior distribution for* $Y_{48,1,3}$.

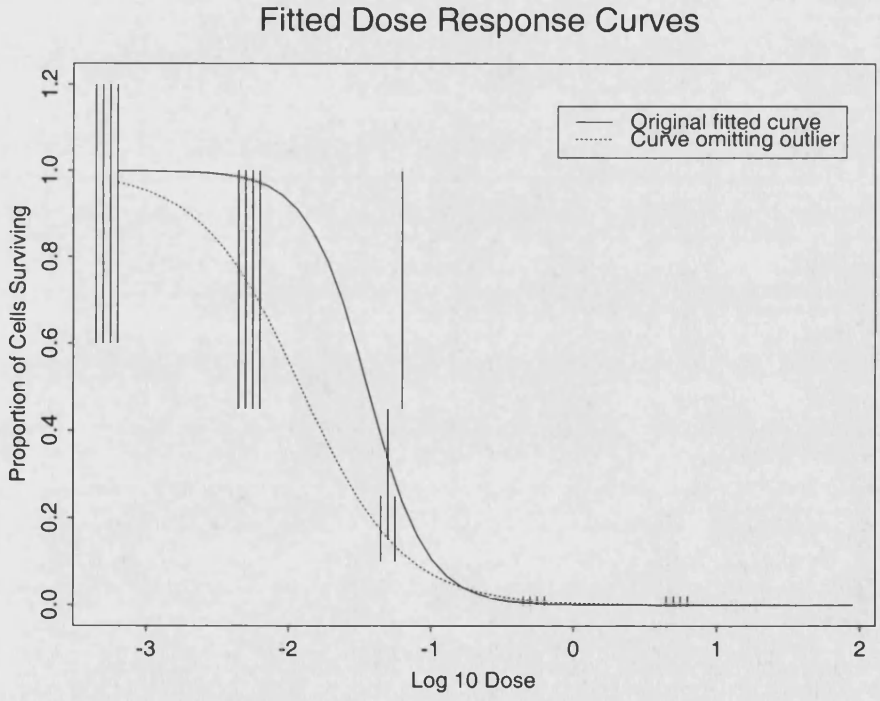

Figure 8-7: *Drug ACD, Patient 48. Proportions of surviving cells.*

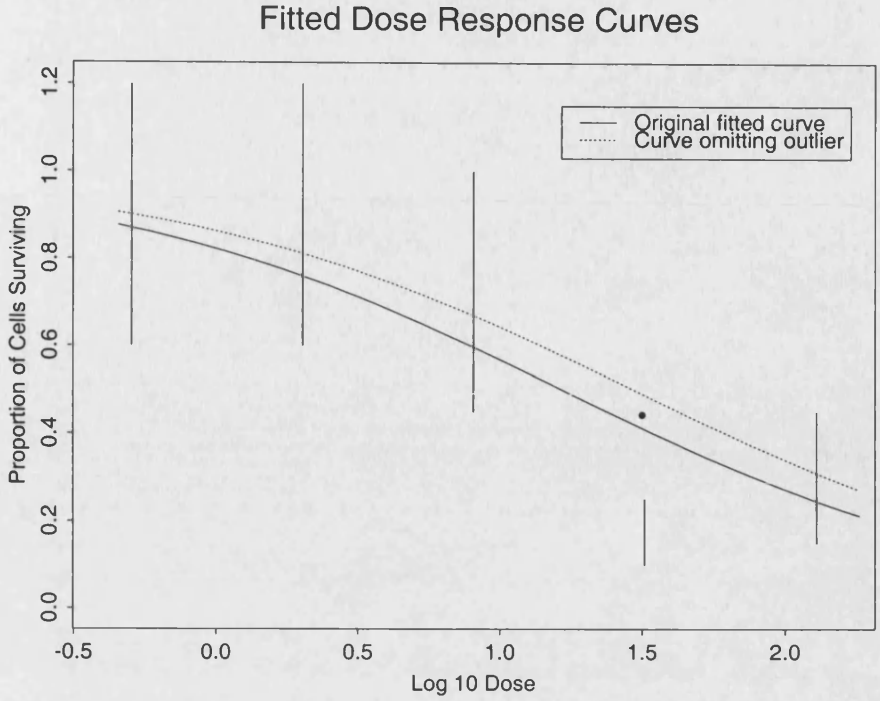

Figure 8-8: *Drug IFN, Patient 48. Fitted Survival Curves.*

the actual dose response curve we cannot say which of the values for doses 4 and 5 has the greater residual. We use the term residual in a general sense since we cannot calculate exact values not having precise data. When we plot the initial dose response curve in Figure 8-8, the initial estimated curve, shown by the solid line indicates it is value for dose 4 which appears to be atypical so it is this value we omit from our second analysis.

The dotted line in this Figure (8-8) shows a dose response curve above the original which is to be expected. The estimated proportion of surviving cells at this dose is 0.44 which is represented on the graph by the point.

The corresponding numerical estimates show a reduction in variation for both slope and LC90. The estimated LC90 has increased slightly with the related percentiles showing an increase in the median but a smaller overall range.

When we examine the posterior distribution of $Y_{48,1,4}$, as depicted in Figure 8-9, we note that the recorded value only appear in 1.48% of the simulations.

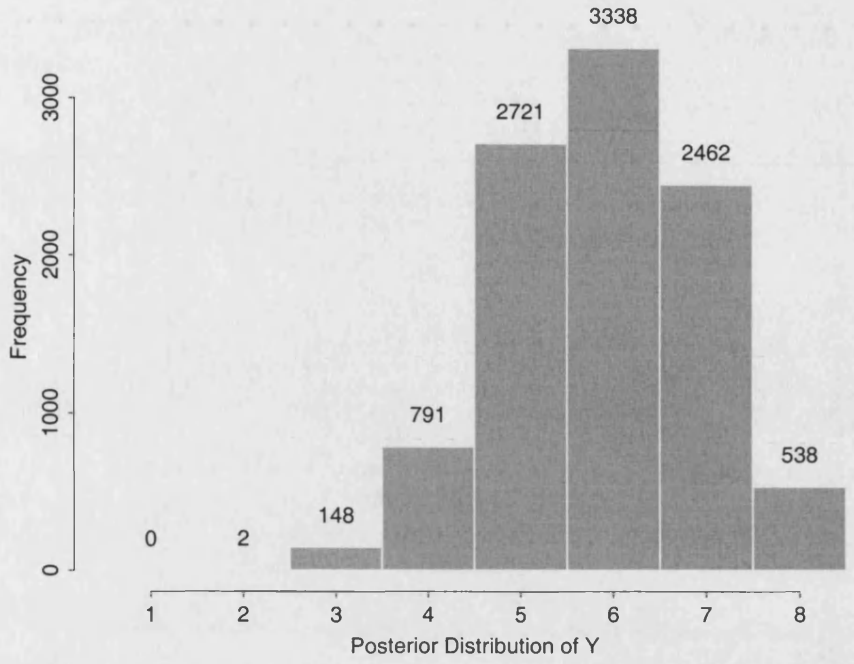

Figure 8-9: *Drug IFN, Patient 48. Posterior distribution for* $Y_{48,1,4}$.

					Ranks		
	$E(X^*)$	$sdX^*)$	$E(B)$	$sd(B)$	5	50	95
Initial Estimate	2.97	0.735	1.26	0.52	11	28	80
New Estimate	3.23	0.599	1.25	0.43	14	31	60

Table 8.5: *Drug IFN, Patient 48. Parameter estimates both before and after omission of potential outlier.*

Spotting this type of potential outlier before we run the initial analysis is harder as this will depend on the underlying dose response curve which we do not know. Whilst the data indicate there may be a problem, it does not readily indicate which of the values does not fit with the model beforehand.

8.5 Conclusions

Existing literature concentrates on assessing the discordancy of the data from the prior whereas we are more concerned with the posterior. Whilst two main schools

of thought exist over whether to consider outliers to come from distributions other than that of the main data or to assume a heavier tailed distribution, in practice we tend towards more simple parsimonious models. We try to explain why outliers may happen and in this problem it may be one of several reasons: a simple mis-recording of the data, an assay has failed to work correctly, differences in the cell survival assessment and a very small control value which can lead to inaccurate data as there may be no "typical" areas of the slides to examine. Another simple explanation could be that the recording categories overlap more than stated and have longer tails than the simplistic tent shaped probability function as depicted in Figure 8-1 which we have used as our prior.

We have looked at some typical examples here where the treatment of potential outliers has differing effects. If the control value is small, the LC90 may differ considerably when simulated. A larger control value leads to a more robust estimate.

In nearly all cases where there has been a reduction in estimated slope, there has also been a reduction in the associated posterior standard deviation. Where the initial analysis had given a high slope and standard deviation the new estimate has both a lower slope and a far smaller standard deviation.

If a value is flagged as a potential outlier because it disagrees with the other recorded values, it is useful to be able to check whether this value has been recorded properly and if not perhaps to perform a recount of surviving cells if possible. As with all outliers, this may be a genuine observation but may be due to some undiscovered problem with the assay procedure.

However, we are confident that these values do not have an important effect on the estimation of the LC90s as long as there is a sufficiently large control value to ensure an accurate estimate. If the control value is small, we have imprecise estimates regardless of the presence or absence of outliers. If the estimated LC90 has a large associated error due to small control values then omitting these rogue values and re-estimating the model will often yield a more precise estimate.

Chapter 9

Conclusions and Further Work

9.1 Clinical implication

We have met our initial aims in providing an automated mechanism which gives reliable estimates for the LC90.

We can produce efficient estimates of the LC90 for each assay performed for each patient and drug combination. This is accompanied by an associated measure of accuracy which is indicative of the reliability of this estimate. However, what the program is unable to do is indicate whether a drug will be effective *in vivo*, but it does give relative rankings with respect to all the other assays performed for each drug. The decision whether to use a particular drug ultimately depends on clinical judgement but more detailed information is now available to the clinician which enables a better decision to be made. Since it is impossible to give each drug to every patient to check the *ex vivo* results against *in vivo* responses, the results from the assays aid the determination of a potentially effective treatment or treatment combination and vastly reduces the chance of non-efficacious drugs being chosen.

As well as providing estimates of the LC90 and dose response curve, we have summarised the information by drug and tumour type. In providing the mean and posterior intervals for our variables of interest, we can assess the potential efficacy of subsequent assays without having to run a long analysis each time.

In the analysis of this data it can be seen that the inter-drug and inter-patient variation, as expected, is great. Although patients with good responses in some drugs will often tend to have favourable results across all assays, this does not mean that poor responders in some drugs will not respond to any drugs.

For some drugs there is less information available initially and as assays continue to be performed, we can utilise this additional data in two ways: firstly as model check by comparing the original analysis with subsequent results and secondly incorporate the additional data to improve the accuracy of results.

The benefits over the original method of curve fitting and manually reading off the LC90 from the resultant graph are manifold. There is a significant time saving in analysing all drug types together which also makes use of the information from each patient. The analysis also takes into account the ranges of values each recorded category represents whereas the graph fitting just used the mid points. We are able to give an idea of the accuracy of our estimates using the posterior distributions obtained which was impossible using the initial graph fitting methods.

This assay based approach to treatment, whereby a range of drugs is tested *ex vivo* for potential patient sensitivity *in vivo*, can be used in other areas of cancer treatment. Different assays method will be more appropriate for different tumour type, the DiSC assay used here did not perform as well on solid tumours, however, the underlying principle is the same.

Further generalisations of these methods could extend to other types of diseases and drugs: wherever a drug induced resistance occurs, it is especially vital to identify an effective treatment as early as possible. Although these assays will not replace clinical trials, there is legislation governing the conduct, much can learned from them, especially when a drug is in its developmental stages.

Since cancers are notoriously difficult to treat effectively and much of the treatment available is highly toxic, new drugs are not tested on healthy volunteers as happens for many other drug types. Often anti-cancer drugs are treated *in vivo* on patients who have failed to respond to prior treatment and as such have poor prognoses. There are problems inherent in this approach not only because of

the drug induced resistance from previous treatments which makes this particular group of patients even more difficult to treat . However, this does give us a chance to check clinical correlations and a response observed here is worth investigating thoroughly as it is harder to achieve.

Early indication of efficacy in a drug's developmental stages is beneficial to both the pharmaceutical industry and clinicians as it can prevent a waste of resources, as well as patients as these resources can be directed in areas of potential benefit earlier.

9.1.1 Assay counts

From the analysis of this data, we would like more accurate counts to remove some of the inaccuracy which may be caused by very wide survival categories. Owing to the labour intensive methods, it is not practical to perform exact counts for every dose level. However, in light of the results from our analyses, the scientists at the RUH are now performing actual counts where there are few cells, rather than simply categorising the survival categories. At higher cell densities where an exact count is very time consuming, the estimated cell survival is rounded to the nearest 5%. This does not appreciably increase the workload involved.

Also, since scientists can differ in their assessments, *all* sets of control values and counts are recorded.

It is too early to tell how big an impact this will have in the estimation of the LC90s but we are confident it will reduce some variation.

9.2 Model review

We shall look at the model implemented and review the methods used and consider alternative methods that could also be implemented where appropriate.

9.2.1 Model specification

We are satisfied that the model in final use, as described in Chapter 4, is a good representation of the data collection process. The inclusion of hyper priors for the parameters of the parent nodes' prior distributions allows these variables more freedom and less dependence than if a straightforward prior distribution had been specified. It is useful to be able to feed back information gleaned from the data by using fairly informative priors based on information given and scientific judgement rather than simply use non-informative or improper priors. Thus the prior distributions in use reflect any knowledge gained with respect to the variables of interest. Other distributions could perhaps have been used as priors but we are satisfied the ones used give acceptable results and make use of the knowledge we have.

The distribution of the categories of surviving cells, y_{ijk}, as shown in Figure 8-1, could benefit from having a wider overlap between categories and longer tails in light of the subsequent results. This may reduce the possible problems of irreducibility as all values would then have some probability attached, however small, hence there would be no illegal or incompatible categories. The issue of irreducibility has been resolved in Chapter 6 with a careful identification of starting values which are compatible with the data and the probability model.

Our choice of the "tent"-like probability function used for the y_{ijk}s was based on information given. We subsequently realised that these categories may not be as precise as originally thought owing to various factors: mainly the small numbers of underlying cells $n0_{ij}$.

We also make the assumption of a logistic relationship between $\log_{10}$ dose response and cell survival. We assume all cells survive at zero dose and all cells are killed at a sufficiently high dose. This gives us a convenient function as we have no further information about this survival curve. It could be that this function is not the most appropriate and that different drugs may have very dissimilar curves. Nevertheless, the assumption made seems not unreasonable in the absence of further data.

9.2.2 Model implementation

Although software is available which carries out MCMC analyses, for example we consider the use of BUGS (Spiegelhalter *et al.*, 1994), this is too slow for our needs and hence we develop our own programs. The development of our own software meant we can incorporate all the data features required and tailor the methods chosen to suit our own specific needs.

We chose the Metropolis Hastings algorithm to perform our updating steps since the distributions in use are not standard and hence not conjugate. We could perform Gibbs updates on some of the steps but in the interests of simplicity we have used the same method throughout.

The calculations for the acceptance probabilities have been carefully coded to try and ensure efficiency since some operations are more "expensive" in terms of computer resources than others.

9.2.3 Model convergence and validation

We need to ensure convergence in order to use our results with confidence. Many factors affect convergence and the rate at which it is achieved which we consider in Chapter 6.

Starting values are theoretically unimportant, as the chain will converge eventually regardless of its starting point (Raftery & Lewis, 1992). However, using greatly over dispersed values will hinder convergence. The problem of possible illegal starting points has been mentioned above and is also dealt with in this chapter.

We are also concerned with the rate at which proposals are accepted. High acceptance rates mean that the difference between existing and proposed values is too small and the chain moves slowly. Conversely, big jumps are rarely accepted and the chain often does not move and may become stuck. This is discussed in Section 6.3. There are many recommendations of suitable acceptance rates to be found in the literature varying between 15%–80%. In light of these we

are satisfied with achieving acceptance rates of around 40%. Our proposals are generated using $x' = x + \delta$, where x is our existing value x' is the proposed value and $\delta \sim N(0, \sigma^2)$, this term is modified for discrete variables to be uniform over a particular range. The variance of δ (or range) is inspected and may be modified to ensure our results are satisfactory. Problems of ensuring positive values, where appropriate, are dealt with by use of a correction factor.

Of vital importance is actual convergence of the Markov Chains. Although acceptance rates may fall within the desirable range, they do not indicate convergence and this needs to be checked separately. The convergence techniques employed (see Chapter 6) in this project are those which are readily available via a public package - CODA (Best *et al.*, 1995). Although much work has been done in this area, we use them purely as additional diagnostic tools since they are not wholly reliable. We noticed that the different diagnostics can give differing results as regards the recommended length of burn in, whether convergence has been achieved, and the run length required in order to produce reliable posterior estimates. These diagnostics can vary greatly depending on starting points, especially over-dispersed starting values and the length of chain actually used to provide the estimates. Paradoxically, the required chain length recommended can *increase* substantially when using a longer chain to calculate the recommended length. Often, the best and arguably most simple diagnostic tool to use is a visual check of the variable trace. No diagnostic can claim to be fool proof: we do not know if there is a small area of probability away from the main area which has yet to be visited, or, if visited, whether the chain spent the proportionately correct time in the respective area. This is where there is a valid argument for running several chains from different over-dispersed starting points. The need for efficiency is also important, running several chains is wasteful as burn in periods are required for each, which are then discarded. Providing a chain has run for a sufficiently long time, we do need to run more than one chain.

These diagnostic checks provide a useful tool when combined with common sense but must not be relied on completely.

9.2.4 Results

The analyses produce a wealth of information which is presented to the RUH in a format which enables informed decisions to be made about potentially effective treatments. The results also build up a larger picture of the characteristics of the efficacy of different drugs with respect to tumour type. In Chapter 7 we present results from just a few patients and drugs in order to illustrate a variety of responses.

As expected, there is considerable variation both within and between patients, drugs and tumour types. Whilst patients who respond well with one drug also tend to respond well in others we cannot use outcome of one assay to predict another reliably. In the same way, a poor responder in some drugs may have good results from another. This demonstrates the value of these assays in determining potentially efficacious treatments. Table 7.1 shows a summary of the main variables of interest. It demonstrates a huge difference between the LC90s (the x^*) and also how the slope of the cell survival curves varies (b). Whilst there is some variation across drugs of the underlying rate of cells (λ) this is not large; the between patient variation is large however, with solid clump tumours having the smallest values since cells clump together and it is the clumps which are counted not individual cells as this is impossible using current methods.

Different tumour types also affect the response to the assay with solid clumps having the worst prognosis in nearly all of the drugs under investigation. Conversely, the CLL samples tend to have the best responses. The main results of Table 7.1 are further subdivided into sample type as in Table 7.2.

The estimated LC90s for patients vary in their accuracy and this will depend on several factors. If there is a very small control value, counts may be inaccurate and small differences in actual numbers of cells can make large differences in proportions. Other factors to note include when the LC90 is not actually achieved within the drug range tested. In these cases there is very little information to use and results are not reliable. Efforts are made to rectify this and the ranges tested are reviewed and modified if this is a consistent problem. Also, missing data may cause problems and adversely affect accuracy.

9.2.5 Future subjects

We also consider the problem of analysing additional assays without the need to run the whole model with just one extra subject added per drug.

We use the posterior distributions of our main analysis to provide informative prior distributions for our secondary analysis.

The posterior distribution of the nodes of interest, (b, λ, x^*) were first checked for any correlation. Since this check offers no evidence otherwise, we shall consider them as independent. Standard distributions were fitted and checked for fit: having shown satisfactory fits we are confident they are satisfactory for our purpose.

We use these fitted distributions as prior distributions for the new model. We obtain estimated values for the LC90 which can be compared against those obtained from the main analysis showing the relative benefit or otherwise for this particular drug for the additional patient.

Whilst there is no need to rerun the main analysis for each extra patient if sufficient data for each drug and tumour sample combination have been collected to obtain reliable posterior distributions, it would be a useful extra validity check to compare these posterior distributions from both the initial and subsequent analyses. We would expect them to be similar unless the new data were drastically different from the original data set.

9.2.6 Model modification

The model we have described in Chapter 4 was chosen to mimic the data collection process as closely as possible. However, in light of recommendations subsequently made in Chapter 7 with regard to changing the system of cell counting and recording this data, modifications would be needed to incorporate this new system. The cell survival proportions, the (y_{ijk}), had previously only been estimated with varying degrees of accuracy, actual counts of the cells are now sometimes made in order to obtain exact values and provide more accurate

data. This provides us directly with values for both $n0_{ij}$ and n_{ikj}, the numbers of control and surviving cells respectively. The need for a categorical y_{ijk} value is then made redundant as we have an exact value for the proportion of surviving cells and not just a range within which to work.

To cope with this new counting system, if the n_{ijk} have been counted, there is now no need to update this variable as it can now be treated as a data node rather than a stochastic node. The model then would have to be capable of coping with both types of data accordingly. Currently, only additional patients are counted in this way and their analysis is treated as distinct from the main body of data. If, however, this new data were included in the main analysis, these modifications would need to be incorporated.

9.2.7 Outliers

There is a paucity of literature dealing with the problems of outliers with respect to Bayesian analyses. Whilst some authors concentrate on comparing the data with the prior distribution using Bayes factors (e.g. Young & Pettit, 1996) we are concerned with the fit of the posterior distribution and the actual data.

Our approach has focused on identifying possible outliers in two ways. The first involves looking at the agreement or otherwise of the values recorded at each dose across replicates. The second type of outlier comes from our assumption that the proportions of surviving cells *decrease* with increased dose; where an *increase* has been recorded we also investigate these values.

Having identified these suspect values they can be referred back to the scientist who can double check that they are not transcription errors or possibly re-check the slides. If there is no obvious explanation for the apparent outlier we can see how it may affect the results.

We aim to assess the effect of these values by omitting them and treating the results as missing and performing updates by accepting or rejecting new proposed values. The posterior distribution resulting from the MCMC analysis is then compared against the original recorded value. We can see from this comparison

how often this value was visited and gauge how likely this value is under our model.

We see that on the whole, these outlying values have a only small effect on the overall results. The effect is magnified in the cases where there is little data or a very small control value present. Otherwise, the simulation of these values tends to give smaller posterior intervals as this estimated value is pulled towards the existing ones.

Another approach not implemented here could include the down weighting of suspect observations so their influence would be lessened. This would preclude being able to check the posterior distributions if the observations had been omitted and estimated instead.

9.3 Further work

9.3.1 Outliers

The popularity of MCMC methods in recent years has generated a huge interest in the application of these methods. Much work has also been done on the vital questions of burn-in periods and convergence, see, for example Brooks & Roberts (1999a, 1999b) or Cowles & Carlin (1996).

However, although literature exists on the treatment of outliers, no formal methods have been derived to check the fit of the resulting model as happens with, say, generalised linear models, where there are well defined procedures to follow.

Although part of our treatment of possible outlying values has been automated, the comparison of posterior distributions and the original observation have been performed manually. There is no procedure to assess whether the discrepancy observed is of importance and thus it is necessarily subjective in this area.

It would be useful to develop methods further along these lines so that suspect values could be objectively assessed for both extremity and influence. In this way

decisions are removed from the analyst and become less subject to human error. Existing methods used to test probability models can be used as a basis on which to investigate further ways to analyse outliers. The main difference being that we are given a whole posterior distribution rather than a single fitted value to compare against our initial value. Although we could use a summary statistic such as the posterior mean this would ignore the vast majority of the available information.

This again indicates the importance of convergence as there is little point assessing outliers against a model which may not have converged.

9.3.2 Model assumptions

Where we have made the model assumptions, there are often many options available to us, for example different prior distributions or proposal distributions. We have used a simplistic distribution to represent the cell survival categories. A more sophisticated approach would allow any value to fall within any category with some probability using a smoother distribution. This would allow more movement of the chain and negate the need for valid starting values as all values are now allowed.

9.3.3 Model extensions

Whilst it is beyond the scope of this thesis, it should be possible to incorporate further data into the decision making process. Information on an individual patient may have some bearing on the effectiveness of a particular drug such as age, sex, stage of disease, previous treatment history etc. Additional information on the drugs such as potential side effects and possible interactions can also be included. A utility type function using as much data as possible may be developed to further aid decision making about particular treatments.

9.3.4 Model Validation

At the moment, it is impossible to say with complete confidence how well our model performs without further data. Since new counting techniques are now being implemented an additional check can also be made about the relative gain in the accuracy of our estimate.

9.4 Conclusions

We have developed the model used for analysis using as much information as possible in order to obtain estimates for parameters describing the dose response curves. Out of the many options which are available to us in terms of methods used, the choice of prior probability distributions etc., we believe the final model gives a reliable mechanism with which to estimate the LC90s and to provide a basis for analysis of future subjects.

Although we are satisfied as to the convergence of the MCMC algorithm fitting the model, there are no foolproof convergence diagnostic methods and we need to take great care in their application.

We use the posterior distributions obtained as new prior distributions for the new model derived to analyse additional subjects. This gives us informative prior distributions for the secondary analysis and removes the need for hyper-prior distributions.

Outlying observations are only of concern when there is little other data, for example missing data or small control values. Associated posterior intervals for the variables of interest in these cases are large reflecting their uncertainty. Little work has been done in this area and there is no formal methodology which currently exists. The literature does not deal with values departing from posterior distributions only from prior distributions. As, ideally, the posterior should be influenced little by the prior distribution and mostly by the data, this approach is not compatible with this project.

We have recommended a new more accurate way of assessing the actual

proportions of surviving cells which requires counting all cells on a slide rather than guesstimating them. This could mean modifying the main model to incorporate both types of data available. However, this is more time consuming and not performed for every assay.

The results obtained from these analyses can be updated when there is sufficient extra data in order to provide more accurate information when analysing future subjects.

Bibliography

Barnett, V., and Lewis T. (1978) *Outliers in Statistical Data* Chap 1, Chichester: Wiley.

Best, N.G., Cowles, M.K. and Vines S.K. (1995) *CODA Manual version 0.30.* Cambridge: MRC Biostatistics Unit.

Besag, J. and Higdon D. (1999) Bayesian analysis of agricultural field experiments *Journal of the Royal Statistical Society B* **61**, 691–746.

Bird, M.C., Bosanquet, A.G., Forskitt, S. and Gilby, E.D. (1988) Long-term comparison of results of a drug sensitivity assays in vitro with patient response in lymphocytic neoplasms. *Cancer*, **61**,1104–1109.

Bosanquet, A.G. (1991) Correlations between therapeutic response of leukaemias and in vitro drug sensitivity assay. *The Lancet*, **337**, 711–714.

Bosanquet, A.G. (1994) Short-term in vitro drug sensitivity tests for cancer chemotherapy. A summary of correlations of test results with both patient response and survival. *Trend Exp Clin Med.*

Box, G.E.P., (1980) Sampling and Bayes' inference in scientific modelling and robustness. *Journal of the Royal Statistical Society A* 143, 383-430.

Box, G.E.P., Jenkins, G.M., and Reinsel, G.C. (1994) *Time Series Analysis, Forecasting and Control* Chap 2, Englewood Cliffs, New Jersey: Prentice-Hall.

Brockwell, P.J., and Davies, R.A. (1991) *Time Series: Theory and Methods* Chap 4, New York: Springer-Verlag.

Brooks, S.P., and Roberts, G.O. (1999) Diagnosing convergence of Markov Chain Monte Carlo algorithms *Statistical Computing* **8:4**, 319–335.

Brooks, S.P., and Roberts, G.O. (1999b) On quantile estimation and Markov chain Monte Carlo convergence. *Biometrika* **86**:3, 710–717.

Chatfield, C. (1988) *Problem solving: A statisticians guide* London: Chapman & Hall.

Chatfield, C. (1989) *The Analysis of Time Series: An Introduction* Chap 3, London: Chapman & Hall.

Cowles, M.K. and Carlin B.P. (1996) Markov chain Monte Carlo convergence diagnostics: A comparative review *Journal of the American Statistical Association* **91:434**, 883–904.

de Finetti, B. (1961) The Bayesian approach to the rejection of outliers. *Proceeding of the Fourth Berkeley Symposium on Mathematical Statistics and Probability* (ed. Neymann, J.), 199-210.

Gelfand, A.E., Dey, D.K., and Chang, H. (1992) Model determination using predictive distributions and implementation via sampling-based methods (with discussion). In *Bayesian Statistics 4* (eds. Bernardo, J.M., Berger, J.O., Dawid, A.P., and Smith A.F.M.) Oxford: Clarendon Press, 147–168.

Gelfand, A.E., Hills, S.E., Racine-Poon, A., and Smith, A.F.M. (1990) Illustration of Bayesian inferences in normal data models using Gibbs sampling. *Journal of the American Statistical Association* **85**, 972–985.

Gelman, A., Carlin, J.B., Stern, H.S., and Rubin, D.B. (1995) *Bayesian Data Analysis.* London: Chapman & Hall.

Gelman, A., Roberts, G.O., and Gilks, W.R. (1996a) Efficient Metropolis jumping rules. In *Bayesian Statistics 5* (eds J.M. Bernardo, J.O Berger, A.P. Dawid and A.F.M Smith), pp 599–607. Oxford: Oxford University Press.

Gelman, A., Roberts, G.O., and Gilks, W.R. (1996b) Weak convergence and optimal scaling of random walk Metropolis algorithms. *Technical Report.* Cambridge University.

Gelman, A., and Rubin, B.R, (1992) Inference from iterative simulation using multiple sequences. *Statistical Science* **7**, 457–511.

Geman, S., and Geman, D. (1984) Stochastic relaxation, Gibbs distributions, and Bayesian restoration of images. *IEEE Trans. Pattn Anal. Mach. Intell.*, **6**, 721–741.

Geweke, J. (1992) Evaluating the accuracy of sampling-based approaches to the calculation of posterior moments. In *Bayesian Statistics 4* (eds J.M. Bernardo, J.O Berger, A.P. Dawid and A.F.M Smith), pp 169–193. Oxford: Oxford University Press.

Gilks, W.R., Richardson S, and Spiegelhalter D.J. (1996) Introducing Markov Chain Monte Carlo. In *Monte Carlo Markov Chain in Practice* (eds W.R. Gilks, S. Richardson and D.J. Spiegelhalter), pp. 1–19. London: Chapman & Hall.

Green, P.J, and Han, X-L. (1999) Metropolis methods, Gaussian proposals and antithetic variables: Stochastic models, statistical methods and algorithms in image analysis. In *Lecture Notes in Statistics* (eds P. Barone, A. Frigessi and M. Picciori) Berlin: Springer.

Hastings, C, Jr., (1955) *Approximations for digital computers* Princeton N.J.: Princeton Univ. Press.

Hastings, W.K. (1970) Monte Carlo sampling methods using Markov chains and their applications. *Biometrika*, **57**, 97–109.

Hawkins, D.M., (1980) *Identification of outliers* London: Chapman & Hall.

Heidelberger, P., and Welch, P. (1983) Simulation run length control in the presence of an initial transient. *Operations Research.* **31** , 1109–1144.

Kale, B.K., and Sinha, S.K. (1971) Estimation of expected life in the presence of an outlier observation. *Technometrics.* **13**, 755–759.

MacEachern S.N, and Berliner, L.M. (1994) Subsampling the Gibbs Sampler. *The American Statistician.* **48:3**, 188–190.

Metropolis, N., Rosenbluth, A.W., Rosenbluth, M.N., Teller, A.H., and Teller,

E. (1953) Equations of state calculations by fast computing machine. *J. Chem. Phys.*, **21**, 1087–1091.

Müller, P. (1993) Metropolis based posterior integration schemes. *Technical Report.* ISDS, Duke University.

Press, W.H., Flannery, B.P., Teukolsky, S.A., and Vetterling, W.T. (1988) *Numerical Recipes in C.* Cambridge: Cambridge University Press.

Priestly, M. (1981) *Spectral Analysis and Time Series.* London: Academic Press.

Raftery, A.L., and Lewis S.M., (1992a) Comment: One long run with Diagnostics: Implementation strategies for Markov Chain Monte Carlo. *Statistical Science*, **7**, 493–497.

Raftery, A.L., and Lewis S.M., (1992b) How many iterations in the Gibbs sampler? In *Bayesian Statistics 4* (eds J.M. Bernardo, J.O Berger, A.P. Dawid and A.F.M Smith), pp 763–734. Oxford: Oxford University Press.

Ripley, B,D. (1987) *Stochastic Simulation.* Chichester: Wiley.

Roberts, G.O. (1996) Markov chain concepts related to sampling algorithms In *Monte Carlo Markov Chain in Practice* (eds W.R. Gilks, S. Richardson and D.J. Spiegelhalter), pp. 1–19. London: Chapman & Hall.

Roberts, G.O. (1992) Convergence diagnostics of the Gibbs sampler In *Bayesian Statistics 4* (eds J.M. Bernardo, J.O Berger, A.P. Dawid and A.F.M Smith), pp 775–782. Oxford: Oxford University Press.

Silverman, B.W. (1986) *Density Estimation for Statistics and Data Analysis.* Bristol: Chapman and Hall.

Spiegelhalter, D., Thomas, A., Best, N., and Gilks, W. (1994) *BUGS Manual Version 0.30.* Cambridge: MRC Biostatistics Unit.

Tierney, L (1995) Introduction to general state-space Markov chain theory. (1995) In *Monte Carlo Markov Chain in Practice* (eds W.R. Gilks, S. Richardson and D.J. Spiegelhalter), pp. 59-74. London: Chapman & Hall.

Weisenthal, L.M., Marsden, J.A., Dill, P.L., Macaluso, C.K. (1983) A novel dye

exclusion method for testing *in vitro* chemosensitivity of human tumours. *Cancer Res*, **43**, 749–57.

Weisenthal, L.M., Dill, P.L., Kurnick, N.B. and Lippman, M.E. (1983) Comparison of dye exclusion assays with a clonogenic assay in the determination of drug-induced cytotoxicity. *Cancer Res*, **43**, 258–64.

Whittaker, J (1989) *Graphical Models in applied multivariate statistics.* Chichester: Wiley.

Young, K.D.S. and Pettit, L.I. (1996) Measuring discordancy between prior and data. *Journal of the Royal Statistical Society B* **58**, 679–689.